Measurement Error Adjustment in the Offset Variable of a Poisson Model

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon

By
Kangjie Zhang

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics
142 Mcclean Hall, 106 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E6
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Motor vehicle accidents is the main cause of death among teenagers in the US. Car crashes are the leading cause of death among teenagers. The Graduated Driver Licensing (GDL) program is one effective policy for reducing the number of teenage car crashes. Our study focuses on how the GDL program adopted by the state of Michigan in 1997 took effect. We use Poisson regression with spatially dependent random effects to model the county-level teenage car crash counts and consider a measurement error model for the offset as the offset variable is mismeasured. The total teenage population in the county-level is widely used to be a proxy for the teenage driver population when modelling the teenage driver fatality rate. In our case, the data for the teenage driver population are not available in the county-level but the state-level in Michigan. Thus, a measurement error issue arises in the offset variable of our Poisson model, we propose including a measurement error model to account for the difference between the teenage population and teenage driver population. To the best of our knowledge, there is no existing literature to adjust for an offset variable when it is measured with error, and limited research has addressed the measurement errors in the context of spatial data. In this thesis, a Berkson measurement error model with spatial random effects have been applied to adjust the offset variable in a Bayesian framework, and the Bayesian MCMC sampling is implemented in rstan. To check whether the adjustment for the offset variable will bring any differences to our model, we have conducted real data analysis. We found the coefficient of T (time) becomes less significant after the adjustment, which leads to a new finding for the GDL – the reduction number of teen-drivers can help explain the partial effectiveness of this policy.

Acknowledgements

In the first place, I would like to express my sincere gratitude to my supervisors, Dr. Juxin Liu and Dr. Raymond Spiteri, not only for providing the thesis' guidance and financial support, but also for their great patience, encouragement, motivation, and immense knowledge.

Besides my supervisors, I am grateful to Dr. Raymond James Carroll, for his insightful comments and suggestions which broaden my perspectives in research work. I am also very thankful to Yang Liu and Dr. Peng Zhang, who shared their previous work and code with me for my thesis.

My sincere thanks also go to Dr. Shahedul Khan and Dr. Longhai Li, for teaching me statistical courses and answering my every single question from the class.

I would like to thank all of the professors, graduate students, and staffs in the Department of Mathematics and Statistics. I feel very blessed to be in this department and work with so many wonderful people through my MSc study.

I would like to thank my roommate Madeleine Hunter for being a supportive friend and bringing me happiness during my thesis writing. Last but not least, I would like to thank my parents and sister, who have consistently provided me emotional support throughout writing this thesis and my life in general.

Contents

List of Tables

List of Figures

List of Abbreviations

GDL	Graduated Driver's Licensing
CAR	Conditional AutoRegressive
ICAR	Intrinsic Conditional AutoRegressive
ZIP	Zero-Inflated Poisson
MCMC	Markov Chain Monte Carlo

1. Introduction

This chapter contains the following sections. Section 1.1.1 introduces the background of a spatial data case as our motivating data example. Section 1.1.2 displays data sources and descriptions in the state of Michigan. Section 1.2, 1.3 and 1.4 present the literature review of Zero-inflated Data , Measurement Error and Conditional Autoregressive Model. Section 1.5 is the outline of the remaining part of my thesis.

1.1 Motivating Data Example

In this section, I will introduce the motivating example, which is about teenage fatal car crashes that Chen et al. (2014) [1] discussed for the spatial variations in the effectiveness of graduated drivers licensing (GDL) program in the state of Michigan.

1.1.1 Background

The main cause of death among teenagers in the US is from motor vehicle accidents [1]. In many developed countries, teen-drivers have the highest crash involvement rate comparing to other driver age groups [2]. Several approaches to reduce the number of fatal car accidents involving teenagers have been tested on and some are more successful than others. One effective policy that has been widely implemented is Graduated Driver Licensing (GDL) [3]. The three stages of GDL adopted by the state of Michigan in 1997 are as follows:

- stage I: Supervised learner period, novice teenagers are required to drive 20-60 h and hold the supervised license for a minimum length of time (6-12 months).
- stage II: Intermediate stage of GDL, teenagers are allowed to drive independently with restrictions on driving at night and passengers.
- stage III: Full-licensure stage of GDL, grants experienced teenager drivers full driving privileges without restrictions.

Figure 1.1 (reproduced plots and the original ones are in Chen et al. (2014) [1]) is a comparison of teen-driver fatal car crash counts before and after the implementation of GDL. The implementation of GDL in Michigan is year 1997, year 1996 represent the period pre-GDL, and year 2004 represent the period post-GDL. Chen et al. (2014) [1] provided a Conditionally AutoRegressive (CAR) prior to account for spatial dependence among crash counts from adjacent counties. They also concluded that GDL is an effective policy, and it reduced the risk of fatal car crashes among teen drivers in Michigan. In this thesis, we also find the GDL is effective with different modeling strategy. Moreover, we conclude that the drop number of teen-drivers is one of the reasons for the effectiveness of the GDL. Thus, policymakers can refine GDL by reducing the number of licensed teen-drivers.

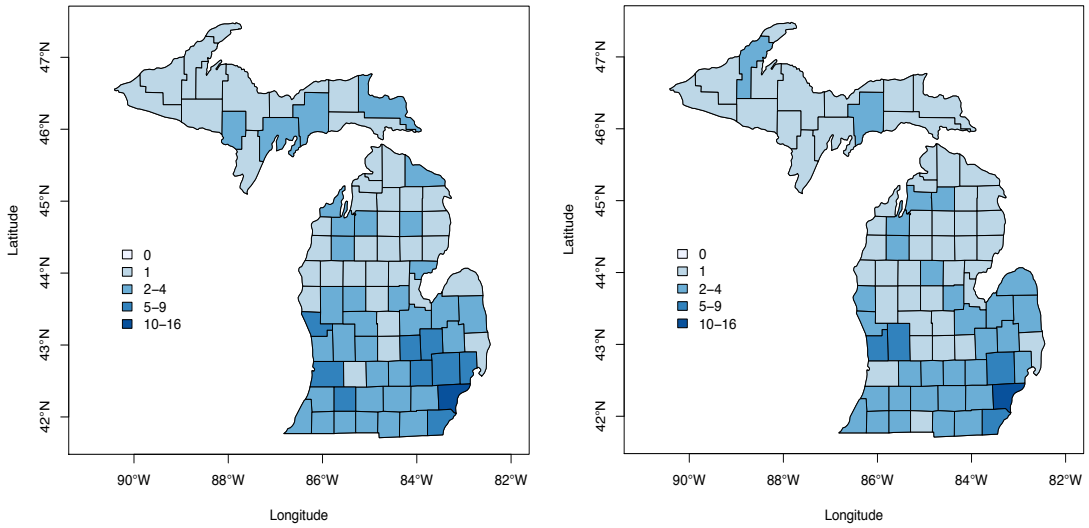

Figure 1.1: Reproduced figure from Chen et al. (2014) [1]: County level teen-driver fatal car crash counts in Michigan in 1996 (left) and 2004 (right).

1.1.2 Data Sources and Descriptions

The data in my thesis is the same as that considered by Chen et al.'s [1] paper. Instead of following Chen et al.'s model, we use a Poisson regression to fit the data. The data used in this thesis are extracted from multiple sources, which are all publicly available. Previous studies showed that unemployment rate [4], Rural-Urban Continuum index [5] were associated with teenager-driver car crash counts. The data for unemployment rate in the state of Michigan

were obtained from the U.S. Bureau of Labor Statistics (https://www.bls.gov/lau/#tables). The Rural-Urban Continuum Codes were downloaded from the US Department of Agriculture (USDA) (https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/).

County level teenager population data in the state of Michigan were extracted from the US Census database. Teenager population is used to adjust yearly counts of fatal teenager-driver crashes for variation in the size of the teenager population across counties, it can be incorporated as an offset term to the model. The population is widely considered as a surrogate for the number of licensed drivers and used for modelling the fatality rate in roadway traffic studies. This leads to a hierarchical model with measurement error, which will be discussed in following sections of my thesis.

Because the recession of automobile industry in Michigan started from 2005 and this lead to changes in driving behavior and the number of car crashes, we followed Chen et al. [1] to collect the data seven years before (1990-1996) and after (1998-2004) the implementation of GDL. In Michigan, the most significant and recent urbanization period occurred during the auto-industrial boom that took place before 1980's. For the period 1990-2004 considered in our study, we can reasonably assume that the Rural-Urban Continuum Codes was constant over time.

Let i index counties in Michigan, $i = 1, \ldots, 83$ and j index years (1990 to 2004 excluding 1997). The data sets consists of the following information:

1. n_{ij}: the number of teenager population in county i and year j ;

2. $X_{1_{ij}}$: unemployment rate in county i and year j ;

3. X_{2_i}: Rural-Urban Continuum Codes in county i, the Rural-Urban Continuum Codes range from 1 to 9 with higher scores indicating more degree of rurality;

4. O_{ij}: the fatal car crash counts for county i at year j.

1.2 Models for Count Data

1.2.1 Zero-inflated Poisson Model

There has been considerable research conducted over the last 30 years focused on predicting car crashes on transportation facilities. Poisson model is one of the commonly used methods for modeling motor vehicle crash data. It is the basis for analyses of rare events, its first applications included descriptions of deaths from mule kicks in the Prussian army [6].

Poisson regression is a member of a family of generalized linear model (GLM) [7] used to model count data and contingency tables. GLM generalizes ordinary least squares (OLS) regression for use with many different types of dependent variables [8]. The GLM family can be applied for binary, ordered categorical, count, and time to failure (or success) dependent variables.

The foundation for Poisson regression is the Poisson distribution, which is a member of the exponential family [8]. Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event [9]. The probability mass function for the Poisson distribution,

$$P(Y = y|m) = \frac{m^y}{y!}e^{-m} \tag{1.1}$$

returns the probability of an observed value y of variable Y which has a Poisson distribution with parameter m. Both mean and variance of a Poisson distribution equal to m. Poisson regression is a GLM with Poisson distribution error structure [8], and it is typically used to model count data.

Poisson regression is a special case of zero-inflated Poisson model without any excess of zero counts. Zero-inflated data are commonly encountered in real world applications, such as traffic accident analysis, in which the accident counts are usually characterized by a large amount of zero observations. In recent years there has been considerable interest in modelling count data with excess zeros. Poisson regression is a common approach to model count data [9]. Quasi-Poisson model and Negative binomial models are also widely used when the count data are overdispersed relative to the Poisson distribution. Modelling

count data with excessive zeros is the extension of above models. One way of capturing both excessive zeros and overdispersion is to use hurdle model, which is the kind of model that is composed by two parts: a point mass at zero and a truncated count distribution (e.g. Poisson distribution) at other non-zero points [10]. Another way is to use a Zero-inflated model [11], which is a mixture of a regular count regression model, such as Poisson or negative binomial model, as well as a component to accommodate the excessive zeros. There are several differences between theses two ways of modelling zero-inflated data. The fundamental difference between hurdle models and zero-inflated models is how the zeros are modelled [11]. Unlike the hurdle models, zero-inflated models make a distinction between structural and sampling zeros. Also, hurdle models can be used to model zero-deflation as well as zero-inflation [12], while zero-inflated models can only apply to zero-inflated data.

The proportion of zero car crash counts for teen-drivers is 51.03% during the period 1990-1996 and 1998-2004 in 83 counties of Michigan. We first try a Zero-inflated Poisson model for our motivating data example because of the following reasons. First, the count data contains more than 50% of zeros. Second, our work is motivated by Chen et al. (2014) [1], in which zero-inflated model were applied to the fatal car crash data in Michigan. Hence, we start with a ZIP model.

Zero-inflated models are based on zero-inflated probability distributions which are able to describe count data sets with excessive zeros. The zero-inflated Poisson (ZIP) model consists of two components to distinguish two different zero generating processes. The first part is a binary distribution that generates structural zeros. The second part is a Poisson distribution that generates counts, some of which may be zeroes and are often interpreted as sampling zeros. The ZIP model defined by [13], can be expressed as follows:

$$O_{ij} \sim \begin{cases} 0, & \text{with probability } \theta_{ij}; \\ \text{Poisson}(m_{ij}), & \text{with probability } 1-\theta_{ij}; \end{cases} \tag{1.2}$$

where m_{ij} is the mean of the Poisson part and θ_{ij} is the probability of zeros for county i and year j observation belonging to excessive zero component. The probability mass function

(PMF) for a ZIP model can be written as:

$$P(O_{ij} = 0|m_{ij}, \theta_{ij}) = \theta_{ij} + (1 - \theta_{ij}) \times e^{-m_{ij}} \tag{1.3}$$

$$P(O_{ij} = k|m_{ij}, \theta_{ij}) = (1 - \theta_{ij}) \frac{e^{-m_{ij}} m_{ij}^k}{k!}, \text{ for } k > 0. \tag{1.4}$$

1. The proportion of extra zeros besides Poisson count model, θ_{ij} is modeled as a logistic regression model, which can be written as:

$$\log(\frac{\theta_{ij}}{1 - \theta_{ij}}) = \beta_{z0} + \beta_{z1} X_{1ij} + \beta_{z2} X_{2i} + \beta_{z3} T_j \tag{1.5}$$

   - T_j is an indicator variable, where $T_j = 0$ if $\text{year}_j < 1997$, $T_j = 1$ if $\text{year}_j > 1997$;
   - β_{z0} represents the intercept of the logistic regression model part;
   - β_{z1}, β_{z2} and β_{z3} represent the coefficients associated with the X_1, X_2 and T in the above zero-inflation part;

2. The mean m_{ij} of the Poisson part can be written as:

$$\log(m_{ij}) = \beta_{m0} + \beta_{m1} X_{1ij} + \beta_{m2} X_{2i} + \beta_{m3} T_j + \log(n_{ij}) + \phi_i, \tag{1.6}$$

   - T_j is an indicator variable, where $T_j = 0$ if $\text{year}_j < 1997$, $T_j = 1$ if $\text{year}_j > 1997$;
   - β_{m0} represents the intercept of the Poisson model;
   - β_{m1}, β_{m2} and β_{m3} represent the coefficients associated with the X_1, X_2 and T in the Poisson model;
   - n_{ij} denotes total number of teenagers in county i and year j in Michigan;
   - ϕ_i denotes county-specific random effects.

When applying the above ZIP model to real data, we got negative intercept and a large absolute value for this intercept, which leads to a very small possibility of extra zeros. Also, the standard errors for parameters in Zero-inflated part are not numerically stable. The estimation results and trace plots for the ZIP model can be found in Appendix A. Therefore, zero inflation part for our motivating data example is not a necessity; we consider a Poisson regression model in the following.

1.2.2 Poisson Regression Model

In this section, we will introduce the Poisson regression model with offset variable included. The link function in Poisson regression uses a logarithmic transformation of the rate that keeps the number of events positive [14], resulting in the following relationship

$$Y_i|\mathbf{X}_i \sim \text{Poisson}\left(m_i\right), \tag{1.7}$$

$$\log\left(\frac{m_i}{N_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} \tag{1.8}$$

where m_i denotes the number of events, N_i represents the number of teenage population of each county in Michigan, and p is the number of predictors (or covariates) in the model.

The expression can be written as follows by moving N_i into the right side of above equation:

$$\log\left(m_i\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \log(N_i) \tag{1.9}$$

The $\log(N_i)$ has a special function in Poisson regression and is usually called the offset [14]. It is needed for some cases when modelling rates is more meaningful than modelling counts.

The offset term in Poisson regression can be chosen by the interests of the research. Different choices for the offset will lead to varying interpretations of the rate. For example, if one is examining the number of aggressive acts in a couple, this value of N_i would be equal to 1.00 for each couple [14]. In person-time analysis, the value of N_i allows for different values in event observation periods across individuals as the basis for estimating event rates. We can also predict fatality rates of car crash problem in a city, where N_i could be the citys population of licensed driver. Michener and Tighe (1992) [15] considered three scale variables as offset when modelling highway fatalities with a Poisson regression: vehicle miles traveled (VMT), registered vehicles, and licensed drivers.

1.3 Measurement Error

There is a critical condition in classical statistical modelling and inference methods: variables included in the models must be measured precisely [16]. However, this condition is

frequently violated in real life setting. Variables measured with error is commonly known as measurement error or errors-in variables problem [17]. One may distinguish misclassification from measurement error if the error-prone variable is discrete. There is a number of books concerning measurement error and misclassification with different focuses. Fuller (1987) [18] summarized the development of measurement error in linear regression models. Carroll (2006) [17] provided comprehensive analysis strategies for regression problems in nonlinear measurement error models, and also discussed some up-to-date methods (e.g. Bayesian analysis, semiparametric and nonparametric methods). Gustafson (2004) [19] provided Bayesian methods which can handle both types of mismeasurement in categorical and continuous variables. Yi (2017) [16] gave a comprehensive overview of topics in modelling and analyzing problems on measurement error and misclassification, and brought miscellaneous methods together in the book.

Relatively few people have addressed the measurement errors in the context of spatial data. Bernadinelli et al. (1997) [20] described Bayesian hierarchical-spatial models for disease mapping with imprecisely observed ecological covariates for spatially correlated data, and posited smoothing priors for both the disease submodel and the covariate submodel. Xia and Carlin (1998) [21] blended methods for spatial-temporal mapping with those for handling errors in covariates in a single hierarchical model framework. Li et al. (2009) [22] quantified the theoretical impact of ignoring measurement error on spatial data analysis in the form of the asymptotic biases in regression coefficients and variance components when measurement error is ignored. They showed that the naive estimators of the regression coefficients are attenuated while variance components are inflated, also showed that biases are related to the spatial dependence parameter. Huque et al. (2014) [23] developed a framework to quantify the bias induced in estimated regression coefficients when covariates are measured with error in spatial regression settings. They extended classical measurement error theory and confirmed the findings of Li et al. (2009) showing that the amount of attenuation depends on the degree of spatial correlation in both the covariate of interest and the assumed random error from the regression model. Huque et al. (2016) [24] proposed a joint modeling approach to assess the relationship between a covariate with measurement error and a spatially correlated outcome in a semiparametric framework. They confirmed that ignoring measurement

error and conducting naive analysis using both generalized additive model and linear mixed model attenuates the estimated regression coefficient toward the null hypothesis of no effect. Spatial data in the presence of covariate measurement errors have not been fully explored. The first novel feature of this thesis is modelling the measurement error with spatial random effects (Conditionaly Autoregressive prior) in a Bayesian framework. Existing work mainly concentrates on the measurement error in covariates and response variable. To the best of our knowledge, no work has been conducted to adjust the measurement error in an offset variable of a Poisson model with a spatial structure. As introduced in 1.2, the offset term is a "structural" predictor, and its coefficient is not estimated by the model but is assumed to have the value 1. In this thesis, we use Poisson regression to model the car crash counts, and the offset term in our model is measured with error. The measurement error adjustment for an offset variable is the second noteworthy feature of this thesis.

1.4 Conditional Autoregressive Model

One of the distinctive feature of spatial data is what Anselin [25] refers to as spatial dependence, the propensity for nearby locations to influence each other and to possess similar attributes. That is, observations from units close together are more similar than those relating to units further apart [26].

Conditional autoregressive (CAR) modelling specifications appear to start with Besag [27]. Bayesian methods are the most commonly used to estimate CAR modelling specifications. Over the past decades, modelling spatial data have been a common problem in a variety of statistical applications, including disease mapping [28], geographical association studies [29], image analysis [30] and agricultural field trials [31]. In particular, they are naturally employed with areal unit data either through single-stage or hierarchical models [32]. In the beginning, the focus is directly on the spatial association of the observations. For the following stage, they are introduced through random effects in the mean structure of the data. Bayesian hierarchical models are typically used in such analyses, where any spatial correlation in the disease data is modelled at the second level of the hierarchy by a set of random effects [26]. These effects are most commonly represented by a Conditional Autore-

gressive (CAR) prior distribution, which is a type of Markov random field [26]. Within this general class of CAR priors, Besag et al. [33] proposed the intrinsic and convolution models, as well as an alternative model by introducing an additional spatial correlation parameter proposed by Cressie [34].

In this thesis, we use Cressie's model to map the spatial pattern in fatal car crash counts over a specific region. Suppose we have a random quantity $\boldsymbol{\phi}= (\phi_1, \phi_2, ..., \phi_n)'$ at n real locations, the CAR model is often expressed via full conditional distributions:

$$\phi_i|\phi_j, j \neq i \sim N(\alpha \sum_{j=1}^{n} b_{ij}\phi_j, \tau_i^{-1}), \tag{1.10}$$

where τ_i is a spatially varying precision parameter, and $b_{ii} = 0$

By Brooks Lemma [35], the joint distribution of $\boldsymbol{\phi}$ is:

$$\boldsymbol{\phi} \sim N(0, [D_\tau(I - \alpha B)]^{-1}). \tag{1.11}$$

If we assume the following:

- $D = \text{diag}(m_i)$: an $n \times n$ diagonal matrix with $m_i =$ the number of neighbors for location i;
- $D_\tau = \tau D$, τ is a precision parameter;
- I: an $n \times n$ identity matrix;
- α: a parameter that controls spatial dependence ($\alpha = 0$ implies spatial independence, and $\alpha = 1$ collapses to an Intrisnic Conditional Autogressive (IAR) specification);
- $B = D^{-1}W$: the scaled adjacency matrix;
- W: the adjacency matrix ($w_{ii} = 0, w_{ij} = 1$ if i is a neighbor of j, and $w_{ij} = 0$ otherwise);

The CAR prior specification can be simplified to:

$$\boldsymbol{\phi} \sim N(0, [\tau(D - \alpha W)]^{-1}). \tag{1.12}$$

The α parameter ensures propriety of the joint distribution of ϕ as long as $|\alpha| < 1$. The parameter alpha is often considered to be between 0 and 1 because positive spatial dependence

(that is, more similar if closer) is commonly seen in practice. In this thesis, we also conduct the data analysis includes the negative spatial dependence with a uniform prior (unif(-1, 1)) and the results suggest a positive spatial dependence for our motivating data example. The results can be found in the appendix B . When α is taken as 1, it leads to the IAR specification, which creates a singular precision matrix and an improper prior distribution.

1.5 Outline of Thesis

The remaining part of this thesis is organized as follows: In Chapter 2, we will introduce the Poisson model with and without measurement error adjustment. Chapter 3 presents real data analysis in three different spatial dependence settings. The main findings of this thesis, together with the discussion of the possible improvement for future work will be summarized in Chapter 4.

2. Proposed Model

This chapter contains the following sections. Section 2.1 introduces the model from Chen et al.'s paper, and explains why we proposes a Poisson model for our motivating example instead of following Chen et al's model. Section 2.2 refines the model with measurement error adjusted. Section 2.3 carries out inference in a Bayesian framework.

2.1 Model without Measurement Error Adjustment

Chen et al. [1] applied a log transformation to the adjusted observed yearly county-level teen-driver fatalities as their response variable. They assumed that this transformed log fatality rates y_{ij} for counties $i = 1, \ldots, I$ in years $j = 1, \ldots, J$ followed a normal distribution with mean μ_{ij} and variance σ^2:

$$y_{ij}|\mu_{ij}, \sigma^2 \sim N\left(\mu_{ij}, \sigma^2\right), \quad i = 1, \ldots, I; j = 1, \ldots, J, \tag{2.1}$$

$$\mu_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T1_j + \beta_4 T2_j + V_i, \tag{2.2}$$

where X_{1ij} and X_{2i} refer to the unemployment rate and rurality, same notation meanings are shared as introduced in Section 1.1.2. V_i denotes a county specific random effects. The transformed log fatality rates y_{ij}, $T1_j$ and $T2_j$ are defined as follows:

$$y_{ij} = \log\left(\left(O_{ij} + 1\right)/n_{ij}\right), \quad i = 1, \ldots, I; J = 1, \ldots, J, \tag{2.3}$$

$$T1_j = \max\left(1997 - \text{year}_j, 0\right); \quad T2_j = \max\left(0, \text{year}_j - 1997\right).$$

The reason that we do not follow Chen et al.'s model is the lack of constraints for the response variable. The adjusted log fatality rate y_{ij} is supposed to be less than 1 since $(O_{ij} + 1)/n_{ij}$ is a value between 0 and 1. Moreover, there are two more differences between Chen et al.'s model and our proposed model. First, instead of using two time terms, $T1$ and $T2$, we simply use an indicator variable T in the regression model. We also include the

results in Appendix D for using two-time terms – T_1 and T_2 as proposed in Chen et al.'s paper, which will be summarized in Chapter 4. Second, we add one more parameter α for controlling the spatial dependence when applying the Conditional AutoRegressive prior for the county specific random effects to our model, which gives us more flexibility to investigate how the spatial random effects would influence measurement error. In Chen et al.'s paper, they used an Intrinsic AutoRegressive prior which assumes the value of α is 1.

Because half of the fatal car crash counts for teen drivers are zero, we start with a Zero-Inflated Poisson model for our motivating data example. When applying the ZIP model to real data, we got negative intercept and the absolute value of this intercept is large, which leads to a very small possibility of extra zeros. Therefore, we simplify our methods to a Poisson model.

Section 1.1.2 has already introduced the motivating data example for my thesis: the data is collected for 83 counties in the state of Michigan from 1990 to 2004. As it has been defined in Section 1.4, we denote i as the county index and j as the year index. The Poisson model without measurement error adjusted can be specified as follows:

$$O_{ij}|m_{ij} \sim \text{Poisson}(m_{ij}) \tag{2.4}$$

$$\log(m_{ij}) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_j + \log(n_{ij}) + \phi_i, \tag{2.5}$$

- T_j is an indicator variable, where $T_j = 0$ if $\text{year}_j < 1997$, $T_j = 1$ if $\text{year}_j > 1997$;
- β_0 represents the intercept of the above model;
- β_1, β_2 and β_3 represent the coefficient associated with the X_1, X_2 and T;
- n_{ij} denotes total number of teenagers in county i and year j in Michigan, $\log(n_{ij})$ is the offset term for the above model;
- ϕ_i denotes a county specific random effects. We apply a CAR prior as defined in 1.3, $\phi|\alpha, \tau \sim \text{CAR}(\alpha, \tau)$, where τ is the precision parameter (1/variance) and α is a parameter that controls spatial dependence.

2.2 Model with Measurement Error Adjustment

In the above section, we modelled rate in the Poisson regression by adding the offset term $\log(n_{ij})$. However, our interest is to model the fatality rate which should be defined as O_{ij}/D_{ij} rather than O_{ij}/n_{ij}, where D_{ij} is the number of teen drivers. The same issue was raised in Chen et al.'s paper as well because the data of teen drivers in Michigan are not available at the county level from the year 1990 to 2004. This leads to a hierarchical model with measurement errors, which is common in real data analysis. Especially in roadway traffic studies, the population is considered as a surrogate for the number of licensed drivers and widely used for modelling the fatality rate. The measurement error in offset causes biases in estimated regression coefficients.

The data of teen drivers in Michigan is available at the state level from the year 1990 to 2004. For comparison, the state level teen population and the number of licensed teen drivers are plotted in figure 2.1 and 2.2, which show that the teen population and the number of licensed teen drivers have opposite trends after the implementation of GDL.

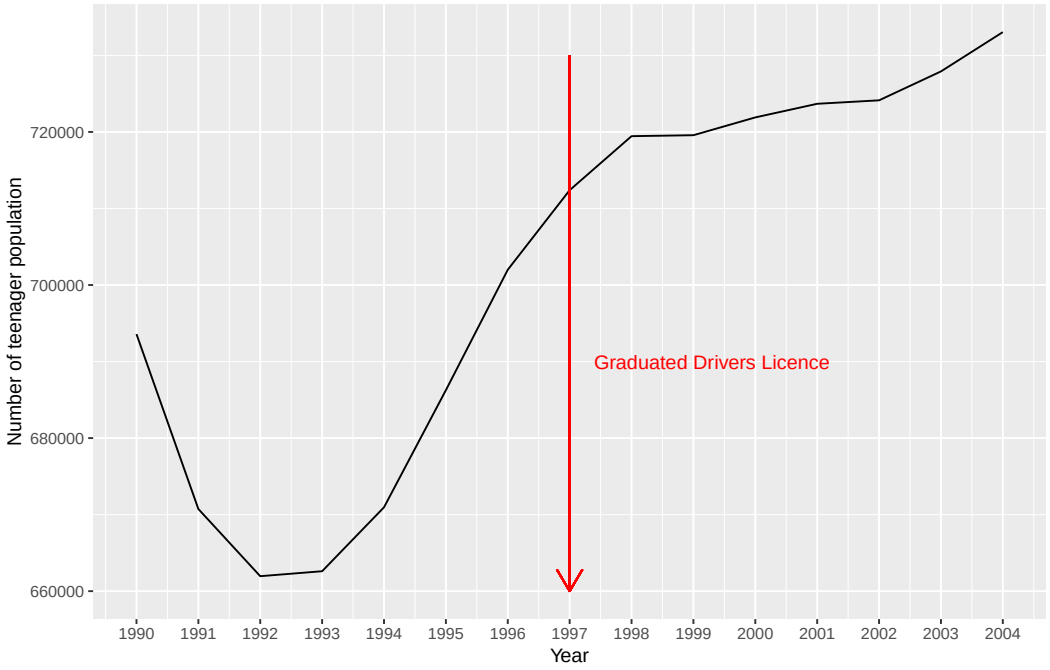

Figure 2.1: State level teen population from year 1990 to 2004

The true offset is suppose to be $\log(D_{ij})$, but the number of licensed teen drivers at the

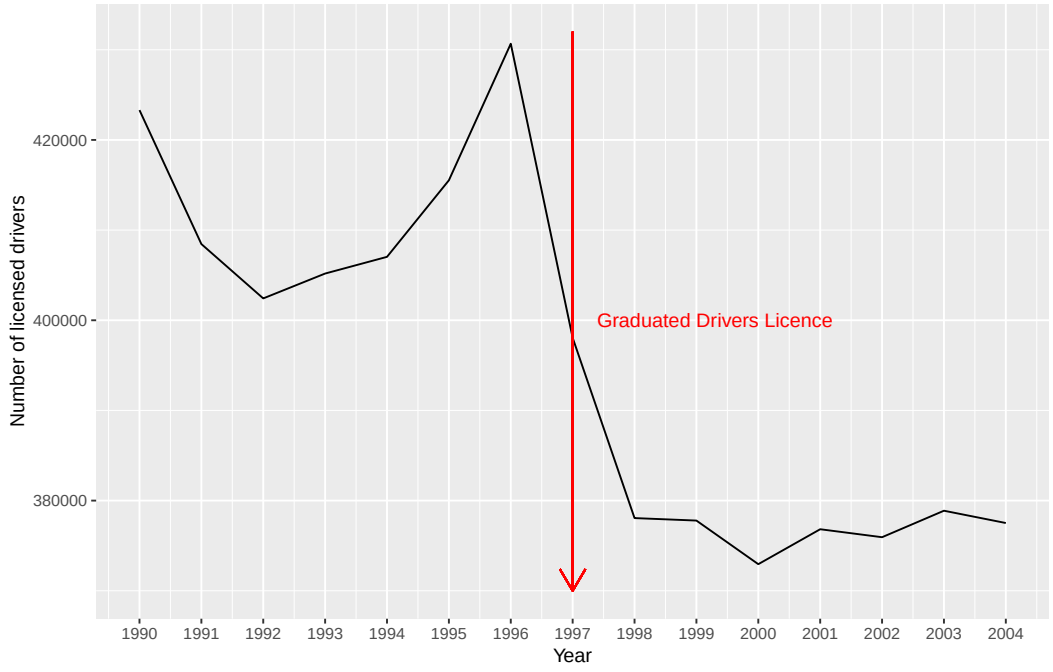

Figure 2.2: State level number of licensed teen driver from year 1990 to 2004

county level are not available. We correct the measurement error in offset by introducing the rate of teen drivers denoted as R_{ij}:

$$R_{ij} = \frac{D_{ij}}{n_{ij}}, \tag{2.6}$$

take the log of both sides of this equation:

$$\log(D_{ij}) = \log(n_{ij}) + \log(R_{ij}), \tag{2.7}$$

- D_{ij} denotes total number of teen drivers in county i and year j in Michigan;
- R_{ij} denotes the rate of teen drivers in county i and year j in Michigan

The Poisson model from section 2.1 with measurement error adjustment could be written as follows:

$$O_{ij}|m^*_{ij} \sim \text{Poisson}(m^*_{ij}), \tag{2.8}$$

$$\log(m^*_{ij}) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_j + \log(n_{ij}) + \log(R_{ij}), \tag{2.9}$$

Because the population of licensed teen driver is obtained at state level, it is natural to think that Berkson measurement error theory should be in operation. The rate of teen

drivers R_{ij} is a value between 0 and 1, thus, we employ a logit transformation to constrain the value for R_{ij}. The logit(p) is defined as the natural log $\ln(p/(1-p))$ where p is a value between 0 and 1. The Berkson measurement error model after logit transformation can be defined as follows:

$$\text{logit}(R_{ij}) = \text{logit}(r_j) + \ \phi_i \tag{2.10}$$

where $r_j = \frac{\sum_{i=1}^{83} D_{ij}}{\sum_{i=1}^{83} n_{ij}}$, represents the rate of teen driver at the state level. Instead of using iid random effects, we apply a CAR prior here in the measurement error model for the county specific random effects ϕ_i. The reason that we use a CAR prior for ϕ_i: 1. spatial dependence can be accounted by the adjusted offset, which is $\log(n_{ij}) + \log(R_{ij})$; 2. adding one more parameter for random effects causes convergence problem and it is not necessary.

2.3 Bayesian Framework

2.3.1 Prior Specification

In general, the prior distribution reflects the prior knowledge or information about the parameter of interest before collecting the data. If there is no such knowledge, then weakly-informative priors need to be assigned to the model parameters.

1. Prior for β: The coefficients of the Poisson regression model will be assigned independent normal priors with zero mean and standard deviation of 100.

    - $\beta \sim$Norm$(0, 100^2)$

2. Prior for $\boldsymbol{\phi}$: We assign a spatial county specific random effects (CAR) for $\boldsymbol{\phi}$= $(\phi_1, \phi_2, ..., \phi_n)$ at $n = 83$ counties in the state of Michigan. We use the same notations as described in section 1.4, $\boldsymbol{\phi} \sim N(0, [\tau(D - \alpha W)]^{-1})$. For simplicity, we can also use the following notation to represent a CAR prior:

    - $\boldsymbol{\phi}|\alpha, \tau \sim$ CAR(α, τ), where τ is the precision parameter (1/variance) and α is a parameter that controls spatial dependence.

3. Prior for $1/\sqrt{\tau}$: We follow Gelman (2006) [36], the variance parameters are assigned uniform $(0, M)$ priors on the standard deviation scale. The commonly used class of inverse gamma (ϵ, ϵ) priors are sensitive to the value of ϵ if the true variance is close to zero, and are therefore not used [26]. A value of $M = 10$ is specified as the upper limit of the uniform prior throughout this thesis.

    - $1/\sqrt{\tau} \sim$ Unif(0, 10)

4. Prior for spatial correlation α: The results of using a uniform prior for α are presented in Appendix C. As we can see the trace plots from C, the parameter α is difficult to estimate from the data alone, so it can be assigned an informative prior.

    - $\alpha \sim$ Beta(γ_1, γ_2), where γ_1 and γ_2 are chosen such that the beta distribution would concentrate around the predetermined value of α.

2.3.2 Software Implementation

Recently, the software package Stan [37, 38] has gained popularity due to its applicability to a broad range of Bayesian models and efficient Markov chain Monte Carlo (MCMC) sampling [37]. Stan differs from BUGS and JAGS in two primary ways. First, Stan is based on a new probabilistic programming language that is more flexible and expressive than the declarative graphical modeling languages underlying BUGS or JAGS, in ways such as declaring variables with types and supporting local variables and conditional statements. Second, Stans Markov chain Monte Carlo (MCMC) techniques are based on Hamiltonian Monte Carlo (HMC), a more efficient and robust sampler than Gibbs sampling or Metropolis-Hastings for models with complex posteriors [37]. In this thesis, we carry out inference in a Bayesian framework, using a Markov Chain Monte Carlo (MCMC) algorithm implemented using rstan.

2.3.3 Posterior Predictive Checks

For model checking, we can generate data from the proposed model and see if the generated data assembles the data we observed. To generate the data used for posterior predictive checks we simulate from the posterior predictive distribution. We can use either the same

predictors that we used for model fitting or new observations of those predictors to simulate data from the posterior predictive distribution. If we use the same values of predictors we denote the resulting simulations by $\boldsymbol{y}^{rep(s)}$, as they can be thought of as replications of the outcome y rather than predictions for future observations [39].

We follow the notation from Gelman et al. (2014) [38] to explain the posterior predictive distribution. Let $\boldsymbol{y}$ be the observed data and $\boldsymbol{\theta}$ be the vector of parameters. To avoid confusion with the observed data, $\boldsymbol{y}$, we define $\boldsymbol{y}^{rep}$ as the replicated data that could have been observed, or, to think predictively, as the data we would see tomorrow if the experiment that produced $\boldsymbol{y}$ today were replicated with the same model and the same value of $\boldsymbol{\theta}$ that produced the observed data. This is the distribution of the outcome variable implied by a model after using the observed data $\boldsymbol{y}$ (a vector of N outcome values) to update our beliefs about unknown model parameters $\boldsymbol{\theta}$. The posterior predictive distribution of $\boldsymbol{y}^{rep}$ can be written as [38]

$$p\left(\boldsymbol{y}^{rep}|\boldsymbol{y}\right) = \int p\left(\boldsymbol{y}^{rep}|\boldsymbol{\theta}\right) p(\boldsymbol{\theta}|y) d\boldsymbol{\theta}. \tag{2.11}$$

Assume S represents the number of posterior samples (warm up samples excluded) and N denotes the number of response variable, for each draw $s = 1, 2, ..., S$ of the parameters from the posterior distribution, $\boldsymbol{\theta}^{(s)} \sim p(\boldsymbol{\theta}|\boldsymbol{y})$, we draw an entire vector of N outcomes $\boldsymbol{y}^{rep(s)}$ from the posterior predictive distribution by simulating from the data model conditional on parameters $\boldsymbol{\theta}^{(s)}$ [39]. Thus, the simulated $\boldsymbol{y}^{rep}$ is an $S \times N$ matrix.

In the Bayesian formulation, the discrepancy between model and data can be measured by the test quantities, $T(y, \theta)$, which depend on both data and unknown (nuisance) parameters by using posterior predictive replications of the data [40]. In our real data analysis part, we use the notation $T(y)$ for a test statistic, which is a test quantity that depends only on data. The tail-area probability for a "test quantity", which is also called the Bayesian p-value, is defined as the probability that the replicated data could be more extreme than the observed data, as measured by the test quantity [38]:

$$p_B = \Pr\left(T\left(\boldsymbol{y}^{\text{rep}}, \boldsymbol{\theta}\right) \geq T(\boldsymbol{y}, \boldsymbol{\theta})|\boldsymbol{y}\right), \tag{2.12}$$

where the probability is found under the joint posterior distribution of replicate data and the

(nuisance) parameters:

$$p_B = \iint I_{T(\boldsymbol{y}^{\text{rep}},\boldsymbol{\theta})\geq T(\boldsymbol{y},\boldsymbol{\theta})} p\left(\boldsymbol{y}^{\text{rep}}|\theta\right) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{y}^{\text{rep}} d\boldsymbol{\theta}, \tag{2.13}$$

which can be estimated from the simulations by $\Sigma_{s=1}^{S} \mathbf{I}_{T\left(\boldsymbol{y}^{\text{rep(s)}},\boldsymbol{\theta}^{(s)}\right)>T\left(\boldsymbol{y},\boldsymbol{\theta}^{(s)}\right)}/S$, where I_A is the indicator function which is 1 if the condition A is true and 0 otherwise [41]. In our real data analysis, the estimation can be simplified as $\Sigma_{s=1}^{S} \mathbf{I}_{T\left(\boldsymbol{y}^{\text{rep(s)}}\right)>T(\boldsymbol{y})}/S$ since we use test statistics $T(y)$ which depend on data only.

Model checking with spatio-temporal modelling on count data can be difficult with classical approach. Posterior predictive checks have been proposed as a Bayesian way to average the results of goodness-of-fit tests in the presence of uncertainty in estimation of the parameters [41]. Its flexibility allows researchers to consider any discrepancy measure of interest.

3. Real Data Analysis

In this chapter, the real data introduced in Chapter 1 will be applied to models proposed in Chapter 2.

3.1 Analysis and Results

3.1.1 Model Estimation

For the model proposed in Chapter 2, we carry out inference in a Bayesian framework with real data and implement in rstan. Stan development team claims that the programs will run faster if the input is standardized to have a zero sample mean and unit sample variance [37]. Thus, we scale the covariate values for X_1, X_2 and T when applying to the stan program, which also improves MCMC convergence. We use two chains, each with 6000 iterations and discard the first half.

The prior distributions for all parameters are specified in section 2.3.1. The Beta distribution is assigned as a prior for α, the spatial dependence parameter in CAR prior for the spatially dependent random effects, which is denoted as $\alpha \sim \text{Beta}(\gamma_1, \gamma_2)$. We conduct the real data analysis under three different priors for α by choosing different values for γ_1 and γ_2. We choose the values for γ_1 and γ_2 such that the prior distribution for α would favor three different strengths of spatial dependence, which include weak (prior1), moderate (prior2) and strong (prior3) spatial dependence.

The posterior mean, standard deviation and 95% credible intervals of parameters in two models are summarized in table 3.1, 3.2 and 3.3 under three different priors for α. Correspondingly, figure 3.1, 3.2 and 3.3 present a comparison of posterior distributions with and without measurement error model included. The results show that the three different choices of prior distribution for α do not affect the estimations.

For a brief reflection, β_1, β_2, β_3 are the regression coefficients associated with the unemployment rate, rurality and time; τ represents a precision parameter of the CAR prior; α is

a parameter for controlling the spatial dependence. The posterior distributions of α in two models are almost identical because an informative prior is assigned for α. We move the CAR prior to the measurement error model from the main model for the adjustment, which causes a big difference for the posterior distribution of τ between two models. The coefficients of the unemployment rate, rurality share the similar posterior distributions in two models, which means the error in the offset of our model is not harmful to these covariates. The posterior distributions of β_0 and β_3 have a relatively noticeable difference between two models.

Table 3.1: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under prior1).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.88	0.06	$(-8.00, -7.75)$	-8.44	0.07	$(-8.57, -8.30)$
β_1	-0.10	0.04	$(-0.19, -0.01)$	-0.10	0.04	$(-0.19, -0.02)$
β_2	0.14	0.06	$(0.02, 0.26)$	0.16	0.06	$(0.03, 0.28)$
β_3	-0.12	0.06	$(-0.24, 0.00)$	-0.30	0.06	$(-0.41, -0.18)$
τ	0.29	0.09	$(0.15, 0.50)$	1.51	0.44	$(0.85, 2.53)$
α	0.20	0.04	$(0.12, 0.29)$	0.20	0.04	$(0.12, 0.29)$

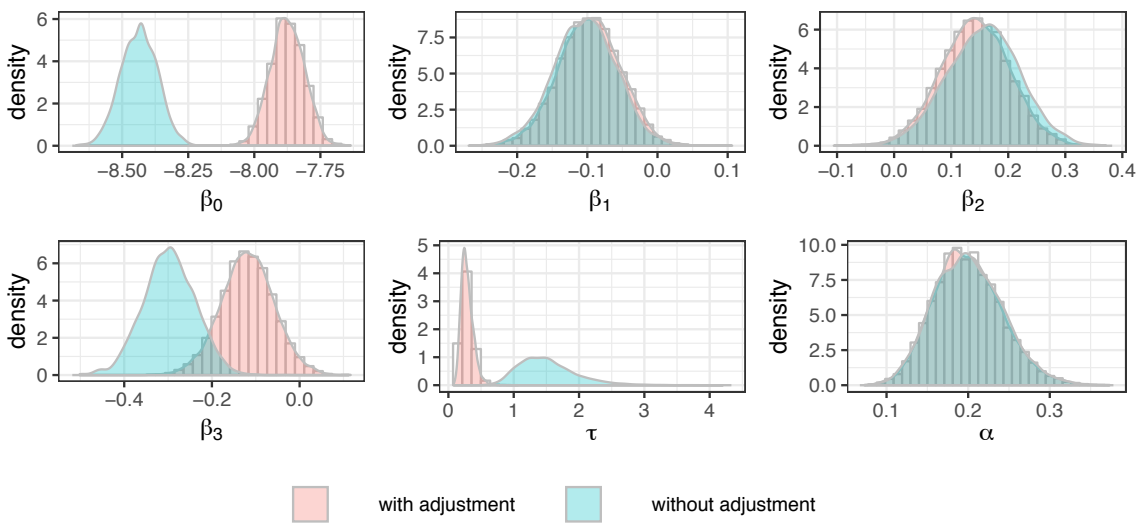

Figure 3.1: Posterior distributions comparisons with and without measurement adjustment (under prior1).

Table 3.2: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under prior2).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.87	0.07	$(-8.00, -7.74)$	-8.44	0.08	$(-8.60, -8.28)$
β_1	-0.10	0.04	$(-0.18, -0.01)$	-0.10	0.04	$(-0.19, -0.02)$
β_2	0.14	0.06	$(0.03, 0.26)$	0.15	0.07	$(0.02, 0.29)$
β_3	-0.12	0.06	$(-0.24, 0.00)$	-0.30	0.06	$(-0.41, -0.18)$
τ	0.29	0.10	$(0.14, 0.51)$	1.54	0.44	$(0.86, 2.60)$
α	0.51	0.05	$(0.41, 0.61)$	0.51	0.05	$(0.41, 0.61)$

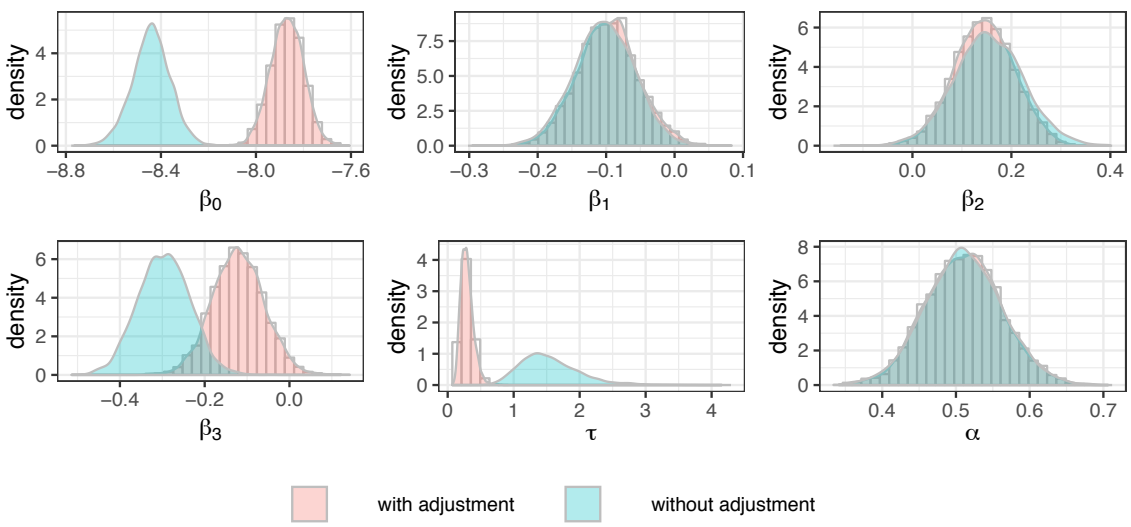

Figure 3.2: Posterior distributions comparisons with and without measurement adjustment (under prior2).

Table 3.3: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under prior3).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.86	0.09	$(-8.03, -7.69)$	-8.44	0.11	$(-8.65, -8.24)$
β_1	-0.10	0.04	$(-0.18, -0.01)$	-0.10	0.05	$(-0.19, -0.01)$
β_2	0.15	0.07	$(0.02, 0.28)$	0.16	0.08	$(0.01, 0.31)$
β_3	-0.12	0.06	$(-0.24, 0.00)$	-0.30	0.06	$(-0.42, -0.18)$
τ	0.34	0.12	$(0.15, 0.61)$	1.70	0.48	$(0.95, 2.81)$
α	0.81	0.05	$(0.72, 0.89)$	0.81	0.05	$(0.71, 0.90)$

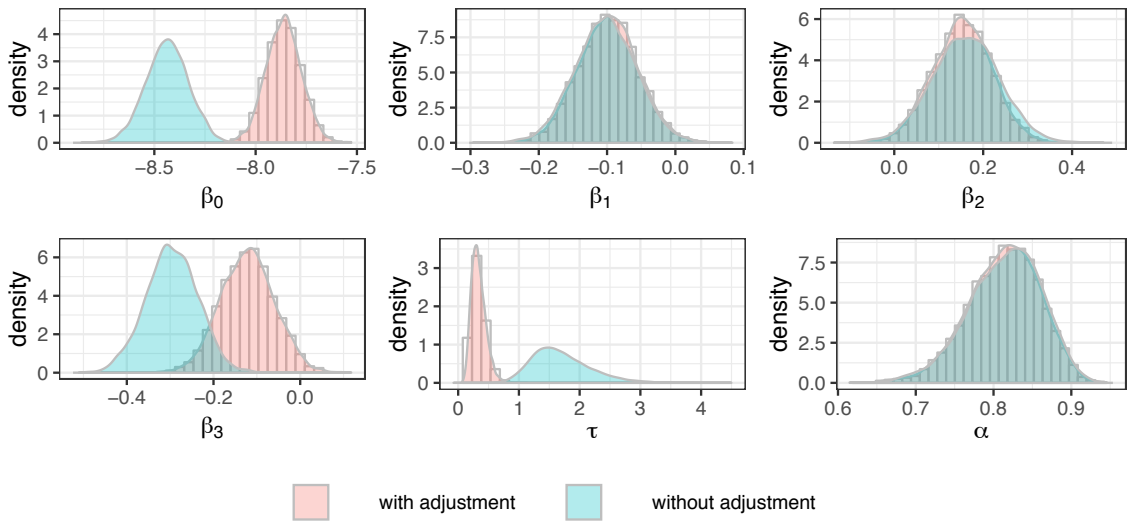

Figure 3.3: Posterior distributions comparisons with and without measurement adjustment (under prior3).

3.1.2 MCMC Diagnostics

Trace plots are used to monitor convergence of the Gibbs sampler, is the realization of the chains versus the iteration number. Figure 3.4 and 3.5 show the trace plots of β_0, β_1, β_2, β_3, τ, and α under the prior2 setting (a prior distribution for α that favors moderate spatial dependence).

To check if 6000 iterations is adequate, we rerun the model five more times on the same data set but with new random initial values for parameters. No matter how we set the initial values for parameters, the chain moves away from its starting values quickly, which suggests that a longer chain is not necessary. No visible patterns are observed; thus, our sampler has mixed well. These trace plots show stationary of the Markov chains. The similar stationary pattern can also be observed under prior1 and prior3 settings. Furthermore, the Gelman-Rubin R statistics [42] is 1 for all posterior estimations, which is also the evidence of stationarity.

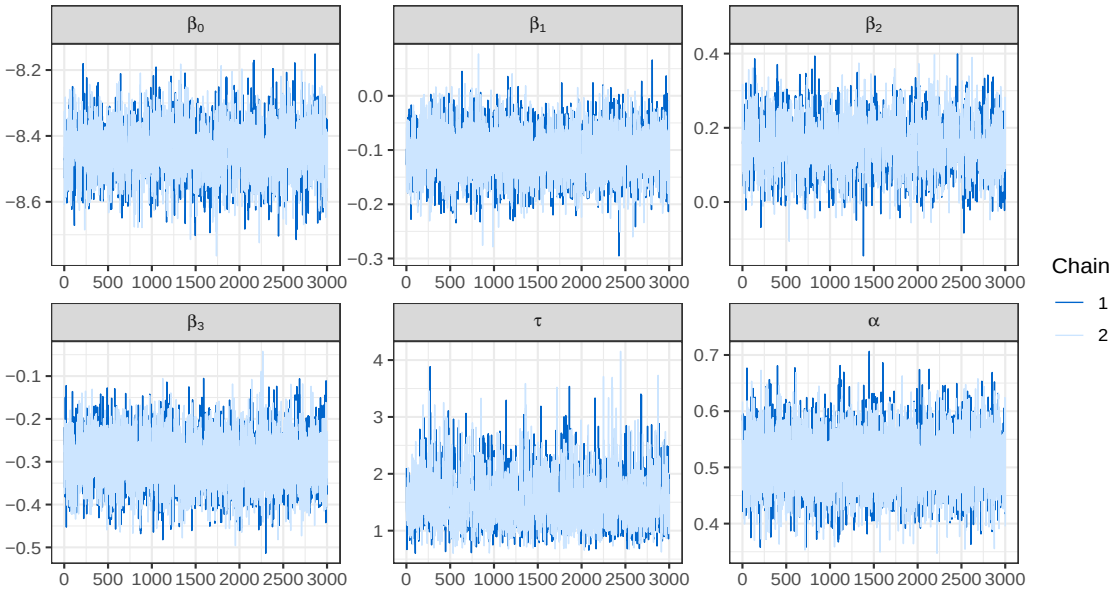

Figure 3.4: Trace Plots for the posteriors of β_0, β_1, β_2, β_3, τ, and α from two Chains without measurement error adjustment (under prior2).

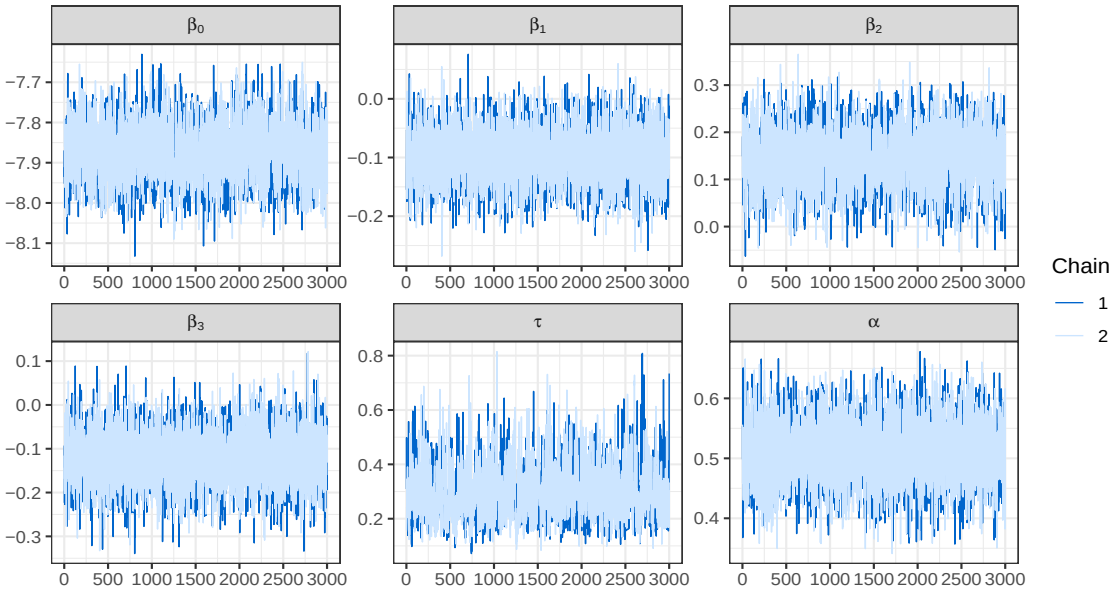

Figure 3.5: Trace Plots for the posteriors of β_0, β_1, β_2, β_3, τ, and α from two Chains with measurement error adjustment (under prior2).

3.2 Model Checking

We check the fit of our model with various aspects of the data using posterior predictive checking and residual diagnostics for the proposed Bayesian hierarchical model. In this section, only the results under the prior2 for α are included, because we get similar results for both posterior predictive checks and residual diagnostics under prior1 and prior3.

3.2.1 Posterior Predictive Checks

We use posterior predictive checking with 6000 replicated data sets generated from posterior predictive distribution. In the following figures, we use red color for the model without measurement error adjustment, and blue for the model with adjustment. In each graph, the darker color represents for the observed outcomes y, and the lighter color standards for the replication data y_{rep} from the posterior predictive distribution.

Figure 3.6 is the histogram comparison of observed y and five simulated data sets from posterior predictive distribution for two models, and x-axis presents the car crash count, y-axis presents the frequency. Similarly, figure 3.6 is the density plot of observed y and all simulated data sets from posterior predictive distribution for two models. These figures show that overall distributions of the observed data and replicated data are similar.

We measure the discrepancy between model and data by defining three test statistics - zero proportion, mean, and max, which are the three aspects of the data we wish to check. Figure 3.8, 3.9 and 3.10 display the posterior predictive distribution of these three test statistics under our model along with the observed value, the test statistic is denoted by $T(y)$. The dark line is at the value $T(y)$, it is the value of the test statistic computed from the observed y. The lighter area is a histogram of the test statistic for 6000 data sets from posterior predictive distribution. These plots make it easy to see that both of the models, with or without measurement error adjusted, can account for the three aspects of the data well.

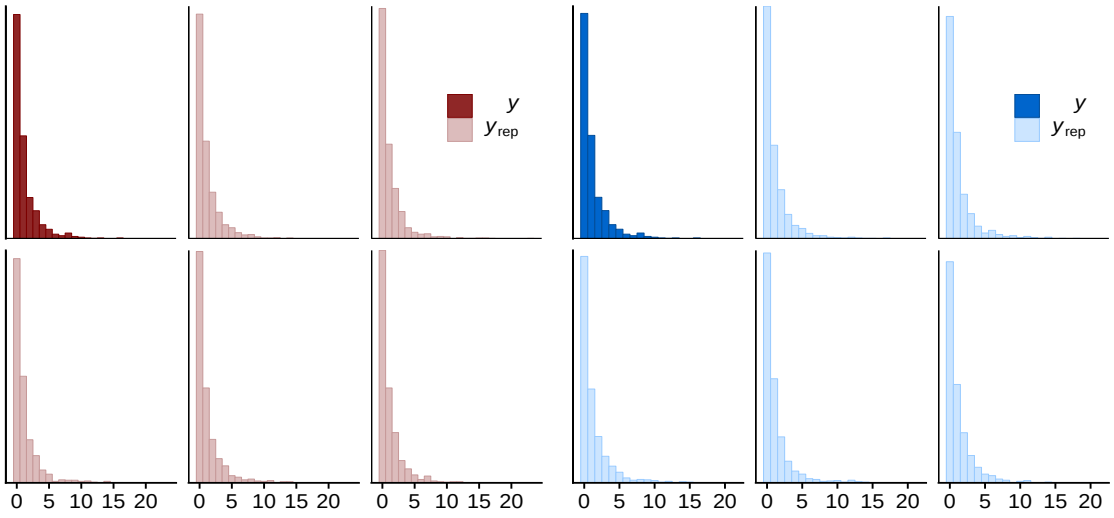

Figure 3.6: Histograms of observed data and replicated data with (right)/without (left) measurement error adjustment (under prior2).

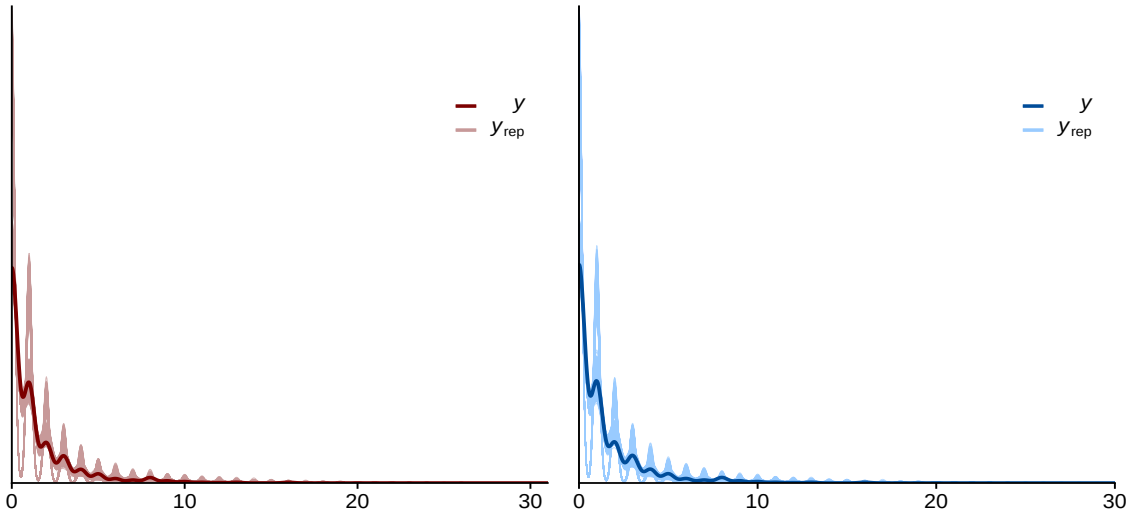

Figure 3.7: Density plot of observed data and replicated data with (right)/without (left) measurement error adjustment (under prior2).

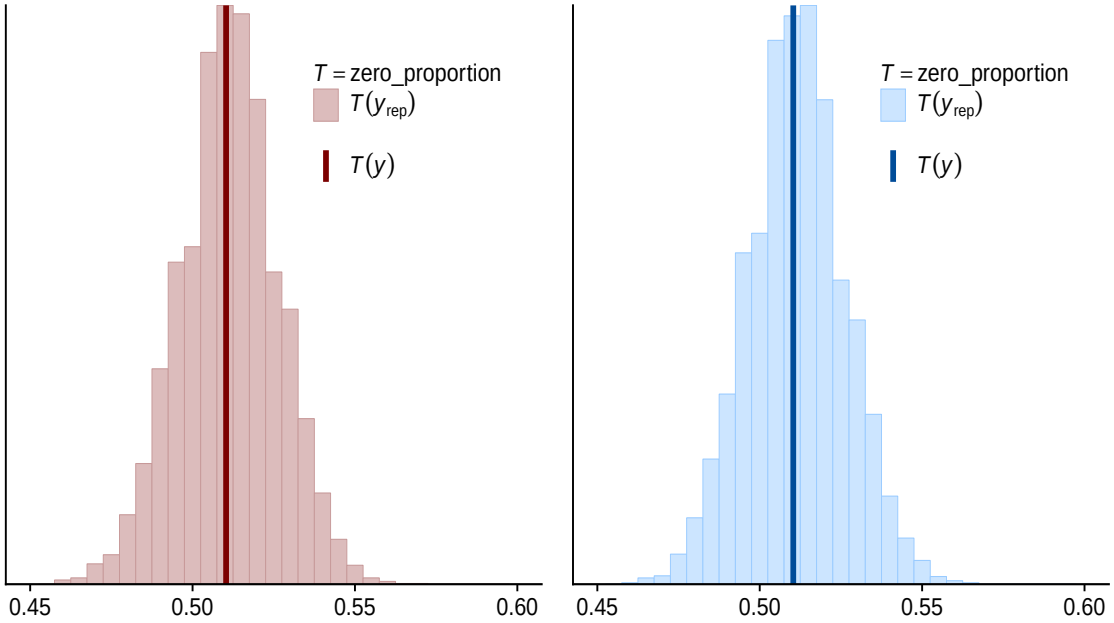

Figure 3.8: Test statistic – zero proportion of observed data and replicated data with (right)/without (left) measurement error adjustment (under prior2).

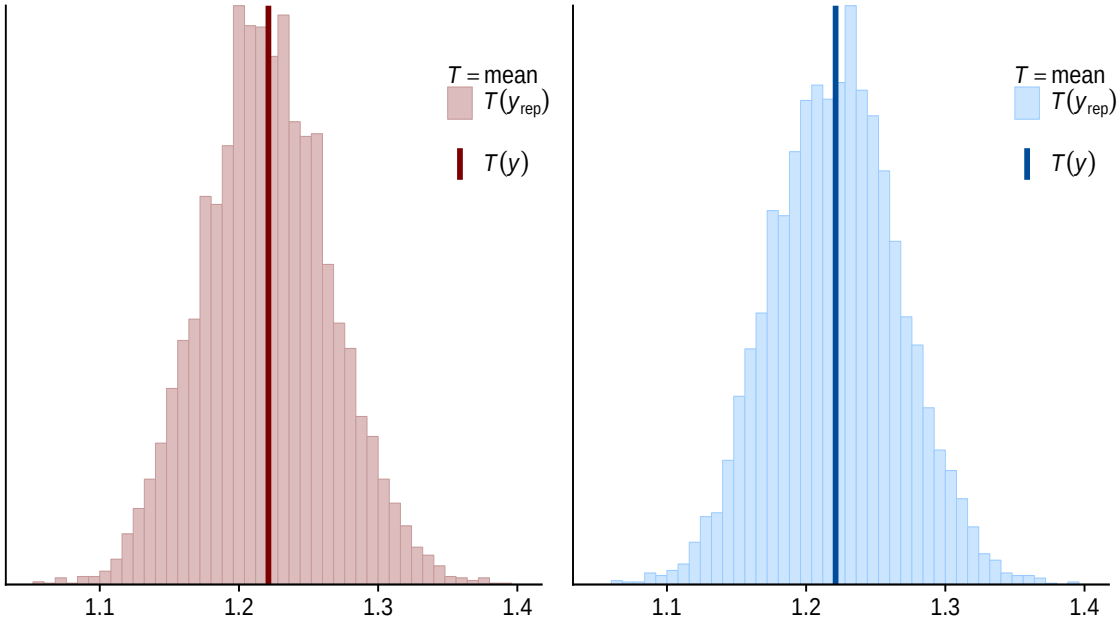

Figure 3.9: Test statistic – mean of observed data and replicated data with (right)/without (left) measurement error adjustment (under prior2).

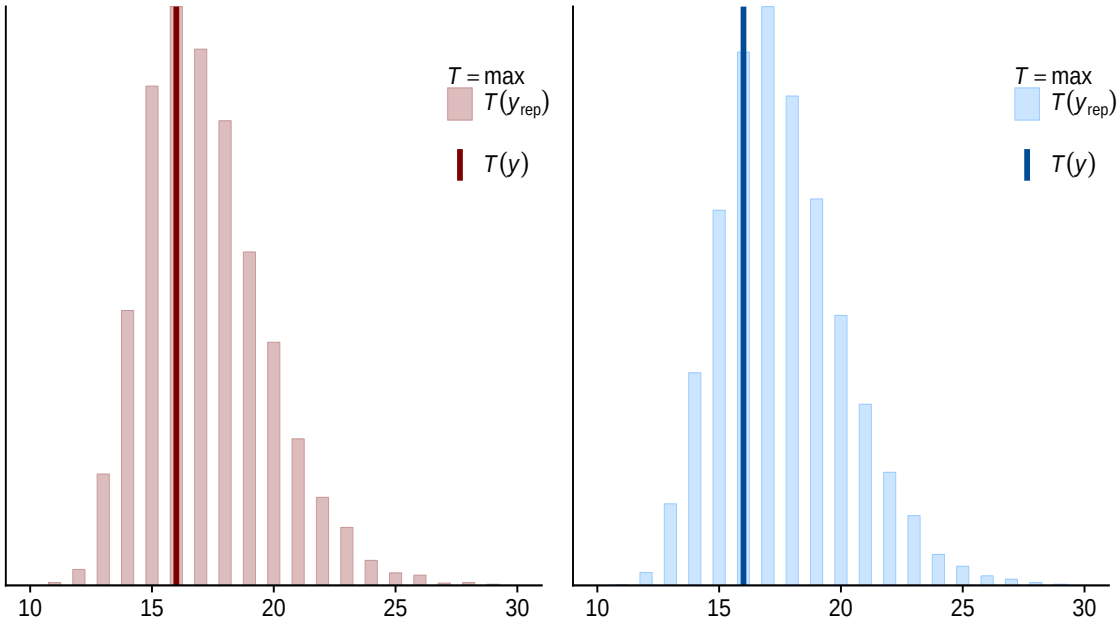

Figure 3.10: Test statistic – maximum of observed data and replicated data with (right)/without (left) measurement error adjustment (under prior2).

3.2.2 Residual Diagnostics

Residuals are commonly used for statistical model checking. In normal linear regression, The residuals are normally distributed when the model is correctly specified. However, for a generalized linear model, the interpretation of standard residual plots can be problematic as the residuals can be non-normality and heteroscedasticity. "DHARMa" [43] is a package to do residual diagnostics for hierarchical regression models and is designed for independent observations only. If a fitted model is correctly specified, then the calculated residuals from "DHARMa" are expected to be uniformly distributed between 0 and 1. In this section, we conduct residual diagnostics by using the package "DHARMa" to create readily interpretable quantile residuals based on the 6000 simulated (replicated) data sets of car crash counts. However, the response variables are not independent in this thesis as we take spatial dependence into our model. This is one of the limitations in this thesis which will be summarized in Chapter 4.

QQ plot is used to check if the DHARMa residuals follow a uniform distribution, it is a graphical method by plotting two probability distributions' quantiles against each other for

distribution comparison [44]. Figure 3.11 and figure 3.12 are QQ plots of DHARMa residuals from two models (without and with measurement error adjustment) under moderate spatial dependence setting. It can be clearly seen that DHARMa residuals from both models are uniformly distributed, which indicate the Poisson model with and without measurement error adjustment are both correctly specified.

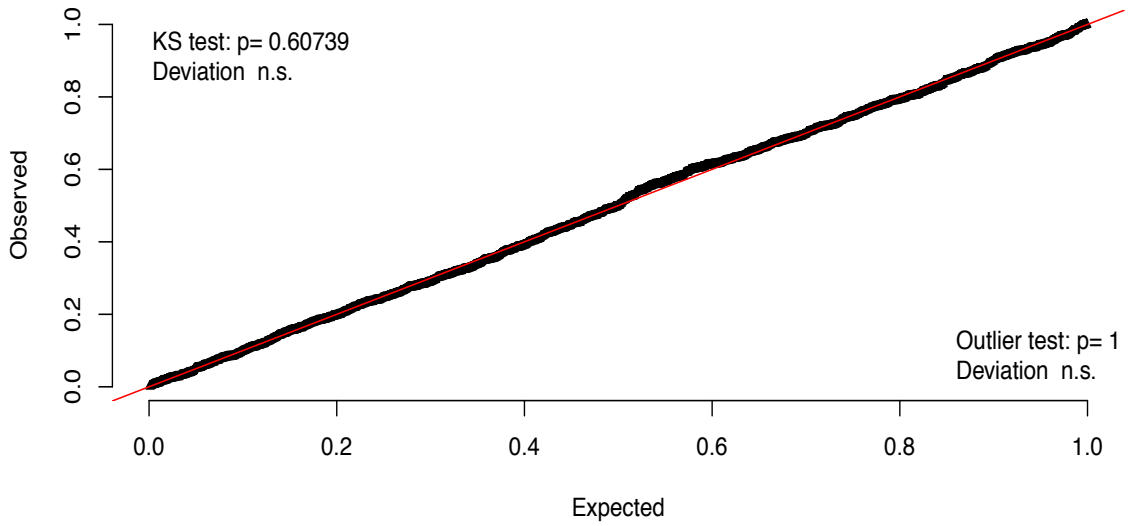

Figure 3.11: QQ plot of DHARMa residuals without measurement error adjustment (under prior2).

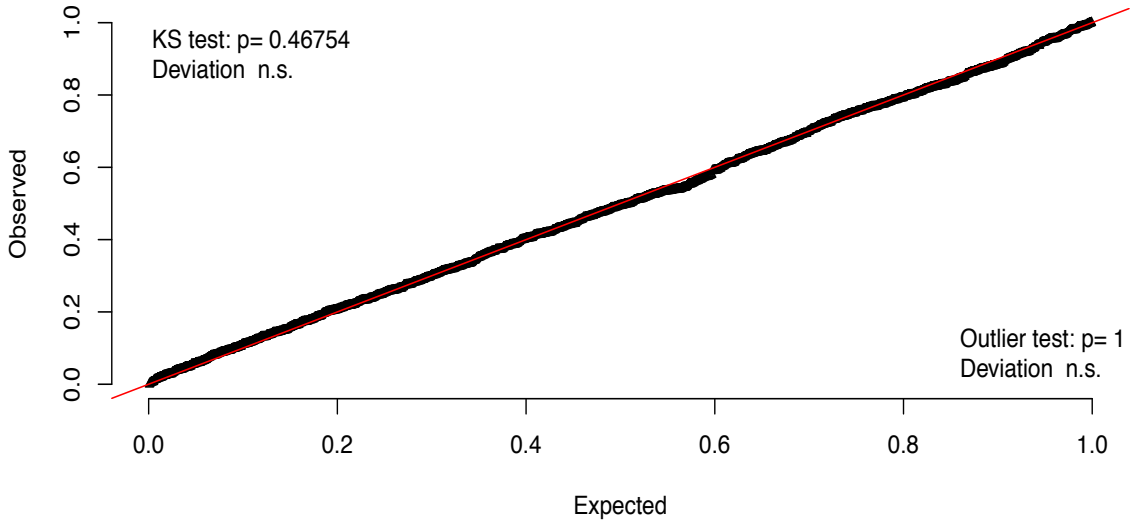

Figure 3.12: QQ plot of DHARMa residuals with measurement error adjustment (under prior2).

4. Conclusion and Future Work

In this thesis, we use Poisson regression to model teen-driver car crash counts with measurement error adjustment in an offset variable and compare the results from the same Poisson model but without the adjustment. We find that both adjusted and non-adjusted model can fit the real data well according to the posterior predictive checks and the residual diagnostics. By comparison, we find that the measurement error in offset term of our model is not harmful to estimating the parameters of X_1 (unemployment rate) and X_2 (rurality). However, the parameter estimations vary in the intercept and the coefficient of T. The coefficient of T is -0.30 without adjustment for the offset. After the measurement error adjustment, the coefficient of T becomes -0.12, which is less significant than before. Similar results can be concluded by using two-time terms in the model from Appendix D. We estimate the difference in time trend before and after GDL implementation via the difference in the two regression coefficients – β_4 and β_3. Under prior2, the posterior mean of $\beta_4 - \beta_3$ is -0.07 with 95% CI $(-0.10, -0.04)$, which suggests a significant decrease in teenage-driver crashes after year 1997 when GDL was implemented before the measurement error adjustment. However, it becomes -0.02 with 95% CI $(-0.05, 0.01)$ after the adjustment, which is less significant than before. We conclude that the reduction of licensed teen-drivers can help explain how GDL took effect in the state of Michigan.

To our best knowledge, there is no existing literature about the adjustment of mismeasured offset in the context of spatial data. In our case, we construct the measurement error model with spatial random effects (CAR prior). In a similar fashion, this adjustment for mismeasured offset can also be applied to other models (e.g. Zero-Inflated Poisson model, Negative Binomial model) in spatial data case.

This thesis has some limitations which can lead to future research work. First, more covariates might be needed for investigating what kind of covariates would be affected by the mismeasured offset. The coefficient of X_1 and X_2 share a very similar posterior distribution before and after the adjustment of measurement error in the offset, while the adjustment

brings different in the estimations of coefficients for variables T. Future research work may include more covariates and quantify the potential effect on parameter estimations. Second, we use the same data for model fitting and checking, so the results from posterior predictive distribution can be overly optimistic. To avoid this overly optimistic problem, out of sample tests (e.g., leave-one-out cross-validation) would be a possible direction of future work. Third, the residual plots from package "DHARMa" is for independent response variables only. The more reasonable residual diagnostics methods and exact statistical justification for our model are still in need of further study. Last, the parameter α for controlling spatial dependence in the CAR prior cannot be 1. However, α is often taken as 1 in the existing literature, which leads to the Intrinsic Conditional AutogRessive (IAR), and it has been widely used in real data analysis to account for the spatial dependence. The reason we use the CAR prior instead of IAR is the parameter α gives more flexibility for accounting spatial dependence, and it can be considered as one of the simulation factors in simulation studies. To the best of our knowledge, measurement error model incorporated with spatial random effects has very limited literature. Therefore, further investigation on whether the strength of spatial dependence will affect the adjustment of mismeasured offset is one possible focus for future work.

References

[1] Yu Chen, Veronica J. Berrocal, C. Raymond Bingham, and Peter X. K. Song. Analysis of spatial variations in the effectiveness of graduated drivers licensing (GDL) program in the state of Michigan. *Spatial and Spatio-temporal Epidemiology*, 8:11–22, April 2014.

[2] Allan F Williams. Teenage drivers: patterns of risk. *Journal of Safety Research*, 34(1):5–15, January 2003.

[3] Jean T. Shope. Graduated driver licensing: Review of evaluation results since 2002. *Journal of Safety Research*, 38(2):165–175, January 2007.

[4] Peter J. Lenk. Bayesian Inference for Semiparametric Regression Using a Fourier Representation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(4):863–879, 1999.

[5] Robert B. Noland and Mohammed A. Quddus. A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention*, 36(6):973–984, November 2004.

[6] Daniel A. Griffith and Robert Haining. Beyond Mule Kicks: The Poisson Distribution in Geographical Analysis. *Geographical Analysis*, 38(2):123–139, April 2006.

[7] Ulf Olsson. *Generalized linear models an applied approach*. Studentlitteratur : [Btj [distributr] ; [ELib [distributr, Lund; [Stockholm, 2006. OCLC: 297807043.

[8] Stefany Coxe, Stephen G. West, and Leona S. Aiken. The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality Assessment*, 91(2):121–136, February 2009.

[9] Frank Avery Haight. *Handbook of the Poisson distribution*. Publications in operations research. Wiley, New York, NY, 1967.

[10] John Mullahy. Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3):341–365, December 1986.

[11] Alain F. Zuur, editor. *Mixed effects models and extensions in ecology with R*. Statistics for biology and health. Springer, New York, NY, 2009. OCLC: ocn288985460.

[12] Brian H. Neelon, A. James O'Malley, and Sharon-Lise T. Normand. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Stat Modelling*, 10(4):421–439, December 2010.

[13] Diane Lambert. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1):1–14, 1992.

[14] David R. Gagnon, Susan DoronLaMarca, Margret Bell, Timothy J. O'Farrell, and Casey T. Taft. Poisson regression for modeling count and frequency outcomes in trauma research. *Journal of Traumatic Stress*, 21(5):448–454, October 2008.

[15] Ron Michener and Carla Tighe. A Poisson Regression Model of Highway Fatalities. *The American Economic Review*, 82(2):452–456, 1992.

[16] Grace Y. Yi. Measurement Error and Misclassification: Introduction. In Grace Y. Yi, editor, *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*, Springer Series in Statistics, pages 43–85. Springer New York, New York, NY, 2017.

[17] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement error in nonlinear models: A modern perspective, second edition*. CRC Press, 2006.

[18] Wayne A. Fuller. *Measurement Error Models*. John Wiley & Sons, July 1987.

[19] Paul Gustafson. *Measurement error and missclassification in statistics and epidemiology: impacts and Bayesian adjustments*. Interdisciplinary statistics. Chapman & Hall/CRC, Boca Raton, 2004.

[20] L. Bernadinelli, C. Pascutto, N. G. Best, and W. R. Gilks. Disease Mapping with Errors in Covariates. *Statistics in Medicine*, 16(7):741–752, 1997.

[21] Hong Xia and Bradley P. Carlin. Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statistics in Medicine*, 17(18):2025–2043, 1998.

[22] Yi Li, Haicheng Tang, and Xihong Lin. Spatial Linear Mixed Models with Covariate Measurement Errors. *Stat Sin*, 19(3):1077–1093, 2009.

[23] Md Hamidul Huque, Howard D. Bondell, and Louise Ryan. On the impact of covariate measurement error on spatial regression modelling. *Environmetrics*, 25(8):560–570, 2014.

[24] Md Hamidul Huque, Howard D. Bondell, Raymond J. Carroll, and Louise M. Ryan. Spatial regression with covariate measurement error: A semiparametric approach. *Biometrics*, 72(3):678–686, September 2016.

[25] Luc Anselin. What Is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis. Technical report, 1989.

[26] Duncan Lee. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, 2(2):79–89, June 2011.

[27] Julian Besag. Statistical Analysis of Non-Lattice Data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24(3):179–195, 1975.

[28] Ying C. MacNab, Andrew Kmetic, Paul Gustafson, and Sam Sheps. An innovative application of Bayesian disease mapping methods to patient safety research: a Canadian adverse medical event study. *Statistics in Medicine*, 25(23):3960–3980, 2006.

[29] D. Lee, C. Ferguson, and R. Mitchell. Air pollution and health in Scotland: a multicity study. *Biostatistics*, 10(3):409–423, July 2009.

[30] Rafael Molina, Aggelos K. Katsaggelos, and Javier Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 8(2):231–246, 1999.

[31] Julian Besag and David Higdon. Bayesian Analysis of Agricultural Field Experiments. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(4):691–746, 1999.

[32] A. E. Gelfand. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–15, January 2003.

[33] Julian Besag, Jeremy York, and Annie Molli. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, March 1991.

[34] Noel Cressie. Regional Mapping of Incidence Rates Using Spatial Bayesian Models. *Medical Care*, 31(5):YS60–YS65, 1993.

[35] D. Brook. On the Distinction Between the Conditional Probability and the Joint Probability Approaches in the Specification of Nearest-Neighbour Systems. *Biometrika*, 51(3/4):481–483, 1964.

[36] Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.*, 1(3):515–534, September 2006.

[37] Stan Development Team. *Stan Reference Manual*.

[38] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. Texts in statistical science. Third edition, 2014.

[39] Jonah Gabry and Tristan Mahr. bayesplot: Plotting for Bayesian Models, August 2018.

[40] Xiao-Li Meng. Posterior Predictive p-Values. *The Annals of Statistics*, 22(3):1142–1160, 1994.

[41] A. Gelman, Y. Goegebeur, F. Tuerlinckx, and I. Van Mechelen. Diagnostic checks for discrete data regression models using posterior predictive simulations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(2):247–268, 2000.

[42] Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.*, 7(4):457–472, November 1992.

[43] Florian Hartig (Theoretical Ecology, University of Regensburg, Regensburg, and Germany). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models, March 2019.

[44] M. B. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis for the analysis of data. *Biometrika*, 55(1):1–17, March 1968.

Appendix A

Results from ZIP Model

The following posterior summary and trace plots are from the ZIP model introduced in section 1.2.1.

Table A.1: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with the ZIP model.

	Poisson Component				Zero-inflated Component		
	Mean	SD	95%CI		Mean	SD	95%CI
β_{m0}	-8.58	0.07	$(-8.73, -8.44)$	β_{z0}	-23.93	5.44	$(-35.86, -14.82)$
β_{m1}	-0.08	0.04	$(-0.17, 0.01)$	β_{z1}	-0.46	3.88	$(-8.09, 7.40)$
β_{m2}	0.15	0.07	$(0.01, 0.29)$	β_{z2}	2.52	3.19	$(-3.50, 9.52)$
β_{m3}	-0.13	0.03	$(-0.19, -0.07)$	β_{z3}	0.52	4.02	$(-7.80, 8.27)$
τ	1.55	0.43	$(0.87, 2.55)$	–	–	–	–
α	0.51	0.05	$(0.41, 0.61)$	–	–	–	–

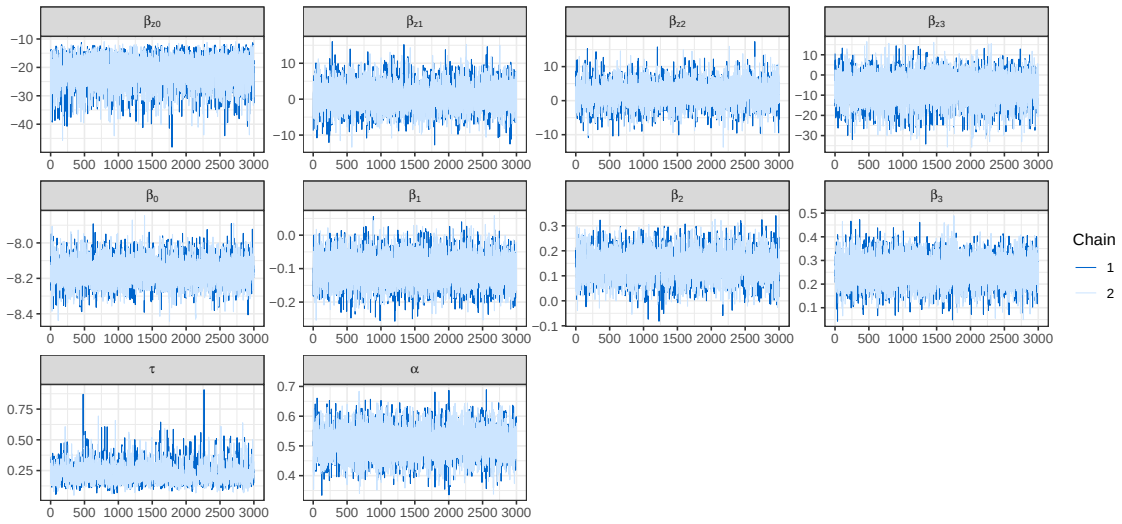

Figure A.1: Trace Plots for the posteriors of parameters in ZIP model.

Appendix B

Results from a Unif(-1, 1) Prior Setting for α

The following posterior summary and trace plots are from the same model settings as model 2.4 and model 2.9 in section 2.1 and 2.2. However, here we use a Unif(-1, 1) prior for α instead of using an informative prior as mentioned in section 2.3.1.

Table B.1: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under a Unif(-1, 1) prior for α).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.87	0.09	$(-8.04, -7.66)$	-8.42	0.14	$(-8.68, -8.10)$
β_1	-0.10	0.04	$(-0.18, -0.01)$	-0.10	0.04	$(-0.19, -0.02)$
β_2	0.15	0.06	$(0.02, 0.27)$	0.16	0.07	$(0.01, 0.31)$
β_3	-0.12	0.06	$(-0.24, 0.00)$	-0.30	0.06	$(-0.42, -0.18)$
τ	0.33	0.13	$(0.14, 0.66)$	1.74	0.54	$(0.95, 3.01)$
α	0.70	0.26	$(0.20, 0.98)$	0.76	0.24	$(0.11, 0.99)$

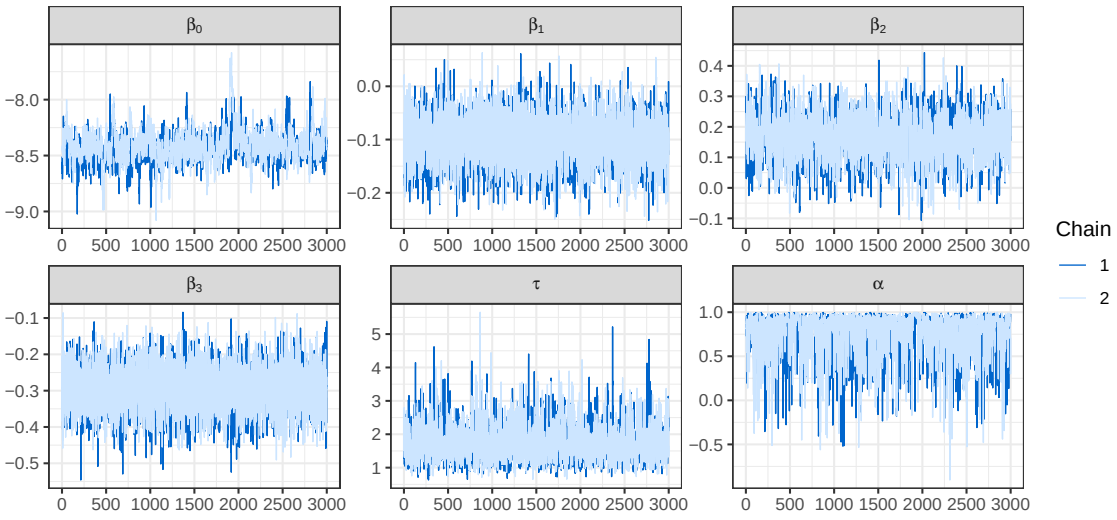

Figure B.1: Trace Plots for the posteriors of parameters without measurement error adjustment under a Unif(-1, 1) prior for α.

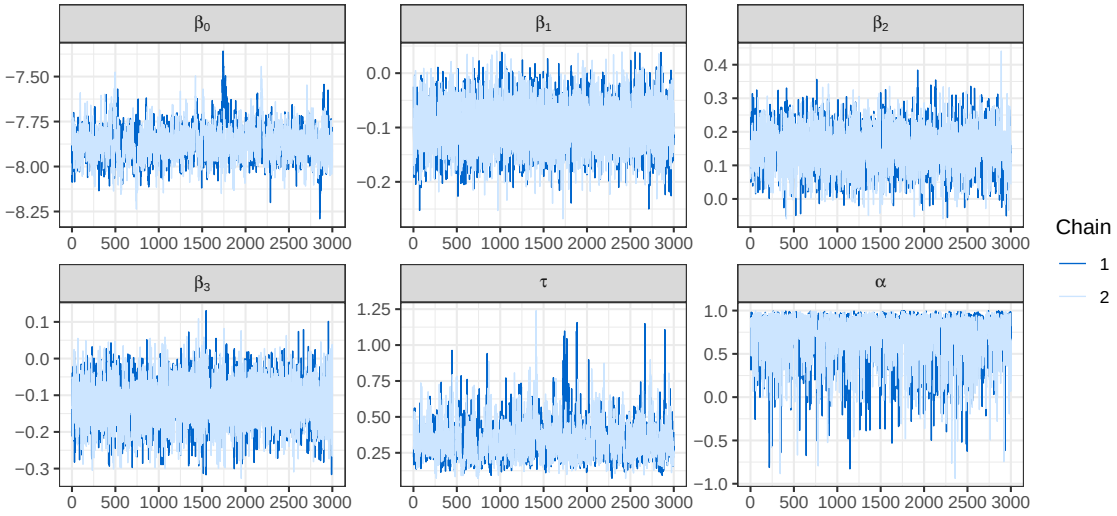

Figure B.2: Trace Plots for the posteriors of parameters with measurement error adjustment under a Unif(-1, 1) prior for α.

Appendix C

Results from a Unif(0, 1) Prior Setting for α

The following posterior summary and trace plots are from the same model settings as model 2.4 and model 2.9 in section 2.1 and 2.2. However, here we use a Unif(0, 1) prior for α instead of using an informative prior as mentioned in section 2.3.1.

Table C.1: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under a Unif(0, 1) prior for α).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.86	0.09	$(-8.03, -7.68)$	-8.43	0.18	$(-8.73, -8.06)$
β_1	-0.10	0.04	$(-0.19, -0.01)$	-0.10	0.05	$(-0.19, -0.01)$
β_2	0.15	0.06	$(0.02, 0.28)$	0.17	0.08	$(0.01, 0.32)$
β_3	-0.12	0.06	$(-0.24, 0.00)$	-0.30	0.06	$(-0.41, -0.18)$
τ	0.34	0.13	$(0.15, 0.66)$	1.74	0.54	$(0.94, 3.03)$
α	0.73	0.21	$(0.20, 0.98)$	0.78	0.20	$(0.22, 0.99)$

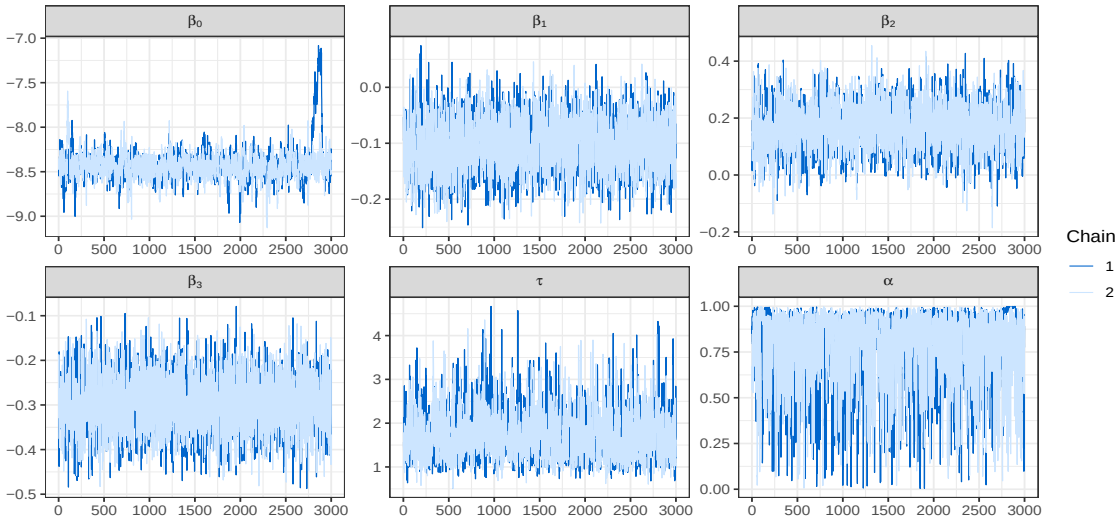

Figure C.1: Trace Plots for the posteriors of parameters without measurement error adjustment under a Unif(0, 1) prior for α.

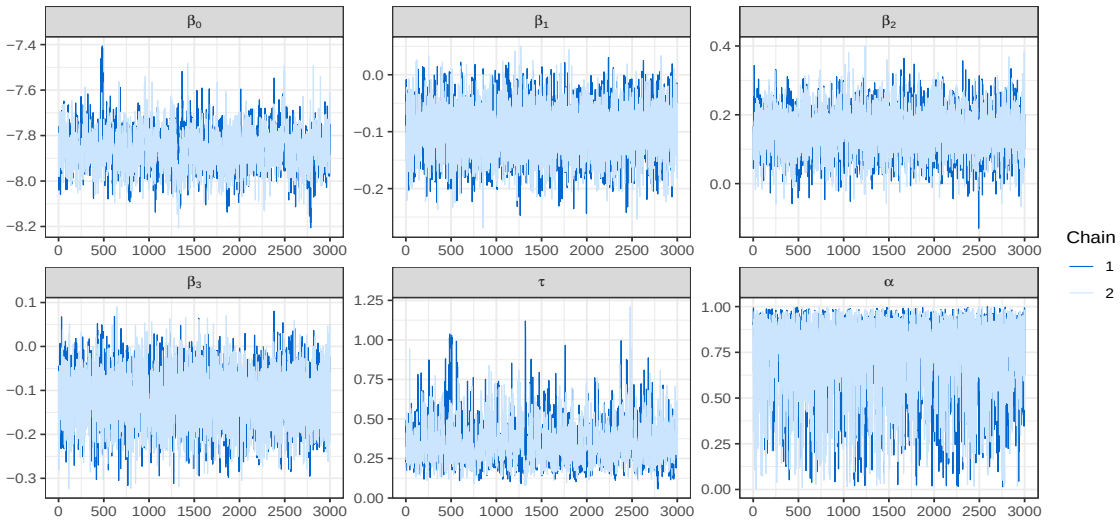

Figure C.2: Trace Plots for the posteriors of parameters with measurement error adjustment under a Unif(0, 1) prior for α.

Appendix D

Results from Model with Two Time Terms

This section presents the results from the same model settings as model 2.4 and model 2.9 in section 2.1 and 2.2 except for using two time terms (T_1 and T_2) in the model.

Model without measurement error adjustment including two time terms:

$$O_{ij}|m_{ij} \sim \text{Poisson}(m_{ij}) \tag{D.1}$$

$$\log(m_{ij}) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_{1j} + \beta_4 T_{2j} + \log(n_{ij}) + \phi_i, \tag{D.2}$$

Model with measurement error adjustment including two time terms:

$$O_{ij}|m^*_{ij} \sim \text{Poisson}(m^*_{ij}), \tag{D.3}$$

$$\log(m^*_{ij}) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_{1j} + \beta_4 T_{2j} + \log(n_{ij}) + \log(R_{ij}), \tag{D.4}$$

$$\text{logit}(R_{ij}) = \text{logit}(r_j) + \ \phi_i \tag{D.5}$$

- $T1_j = \max\big(1997 - \text{year}_j, 0\big)$;
- $T2_j = \max\big(0, \text{year}_j - 1997\big)$;
- β_0 represents the intercept of the above model;
- β_1, β_2, β_3 and β_4 represent the coefficient associated with the X_1, X_2, T_1 and T_2;
- n_{ij} denotes total number of teenagers in county i and year j in Michigan, $\log(n_{ij})$ is the offset term for the above model;
- ϕ_i denotes a county specific random effects. We apply a CAR prior as defined in 1.3, $\phi|\alpha, \tau \sim \text{CAR}(\alpha, \tau)$, where τ is the precision parameter (1/variance) and α is a parameter that controls spatial dependence.

Table D.1: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under prior1).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.92	0.08	$(-8.07, -7.77)$	-8.58	0.08	$(-8.74, -8.42)$
β_1	-0.11	0.05	$(-0.21, 0.00)$	-0.16	0.06	$(-0.28, -0.05)$
β_2	0.14	0.06	$(0.02, 0.26)$	0.18	0.07	$(0.04, 0.30)$
β_3	0.01	0.02	$(-0.02, 0.04)$	0.03	0.02	$(0.00, 0.06)$
β_4	-0.01	0.02	$(-0.04, 0.02)$	-0.04	0.02	$(-0.07, -0.01)$
$\beta_4 - \beta_3$	-0.02	0.02	$(-0.05, 0.01)$	-0.07	0.02	$(-0.10, -0.03)$
τ	0.29	0.10	$(0.15, 0.51)$	1.41	0.40	$(0.80, 2.36)$
α	0.20	0.04	$(0.12, 0.29)$	0.20	0.04	$(0.12, 0.29)$

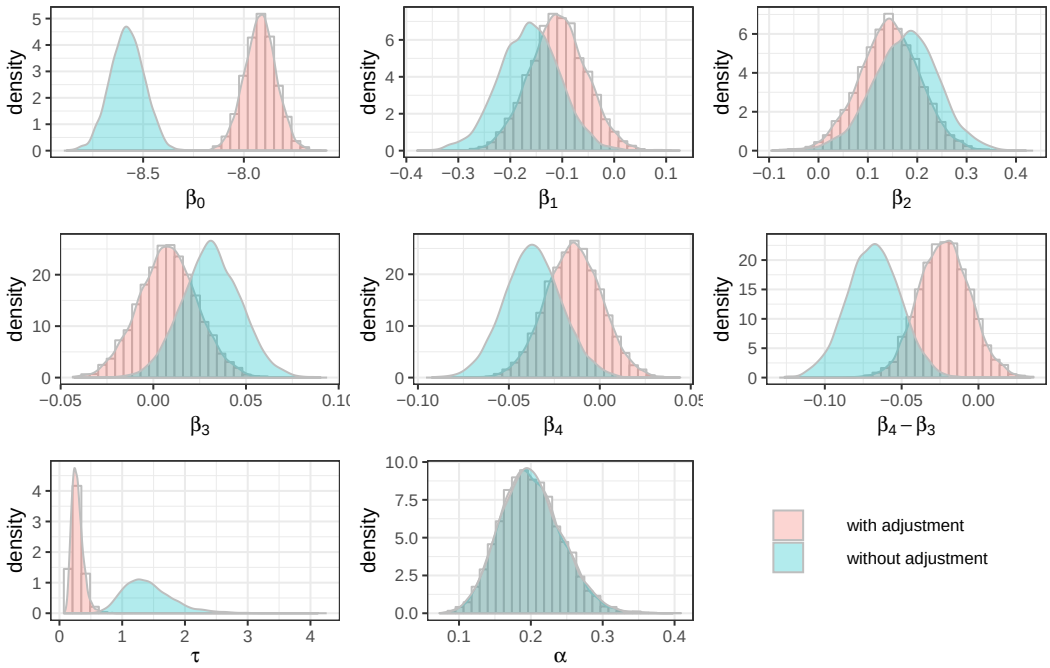

Figure D.1: Posterior distributions comparisons with and without measurement adjustment (under prior1).

Table D.2: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under prior2).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.92	0.08	$(-8.08, -7.75)$	-8.58	0.09	$(-8.77, -8.40)$
β_1	-0.11	0.05	$(-0.22, -0.01)$	-0.16	0.06	$(-0.27, -0.05)$
β_2	0.15	0.06	$(0.02, 0.27)$	0.18	0.07	$(0.04, 0.32)$
β_3	0.01	0.02	$(-0.02, 0.04)$	0.03	0.02	$(0.00, 0.06)$
β_4	-0.01	0.02	$(-0.04, 0.02)$	-0.04	0.02	$(-0.07, -0.01)$
$\beta_4 - \beta_3$	-0.02	0.02	$(-0.05, 0.01)$	-0.07	0.02	$(-0.10, -0.04)$
τ	0.30	0.10	$(0.15, 0.54)$	1.47	0.42	$(0.81, 2.45)$
α	0.51	0.05	$(0.41, 0.61)$	0.51	0.05	$(0.41, 0.61)$

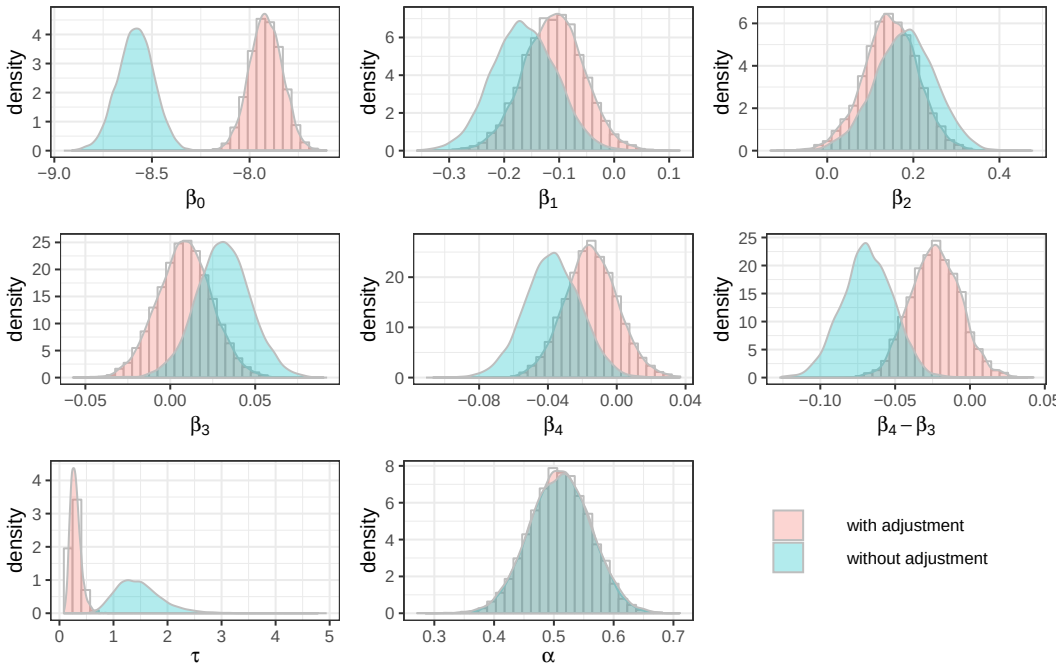

Figure D.2: Posterior distributions comparisons with and without measurement adjustment (under prior2).

Table D.3: Posterior Mean, Standard deviation (SD), and 95% Credible Intervals (CI) of parameters for real car crash data with and without measurement error adjustment (under prior3).

	With adjustment			Without adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-7.90	0.10	$(-8.09, -7.70)$	-8.58	0.12	$(-8.82, -8.35)$
β_1	-0.11	0.06	$(-0.22, 0.00)$	-0.16	0.06	$(-0.27, -0.05)$
β_2	0.15	0.07	$(0.01, 0.28)$	0.18	0.08	$(0.03, 0.34)$
β_3	0.01	0.02	$(-0.02, 0.04)$	0.03	0.02	$(0.00, 0.06)$
β_4	-0.01	0.02	$(-0.05, 0.02)$	-0.04	0.02	$(-0.07, -0.01)$
$\beta_4 - \beta_3$	-0.02	0.02	$(-0.06, 0.01)$	-0.07	0.02	$(-0.10, -0.04)$
τ	0.35	0.12	$(0.17, 0.62)$	1.63	0.46	$(0.92, 2.67)$
α	0.81	0.05	$(0.71, 0.89)$	0.81	0.05	$(0.71, 0.89)$

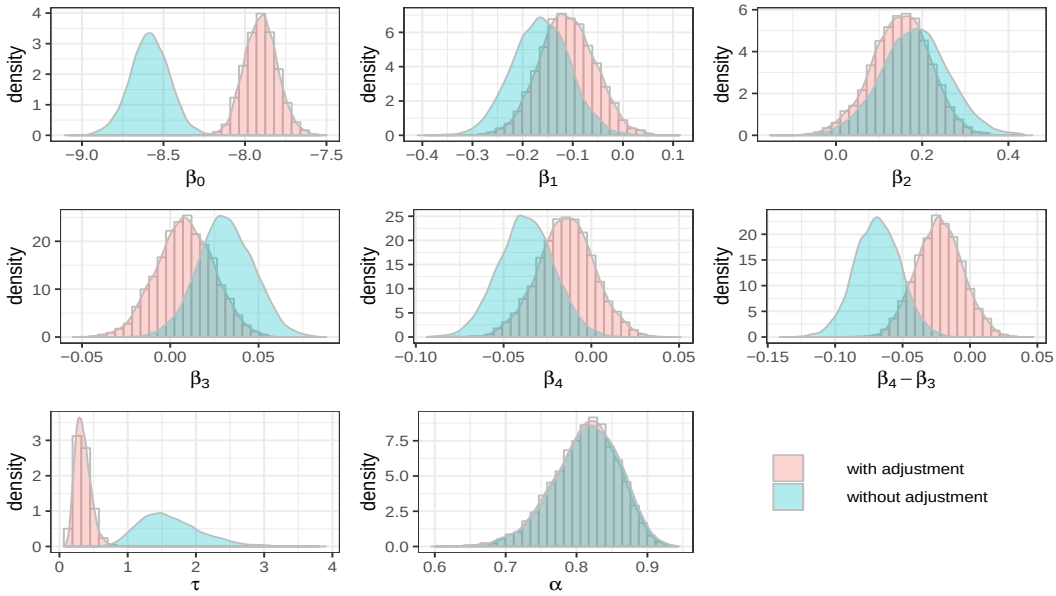

Figure D.3: Posterior distributions comparisons with and without measurement adjustment (under prior3).

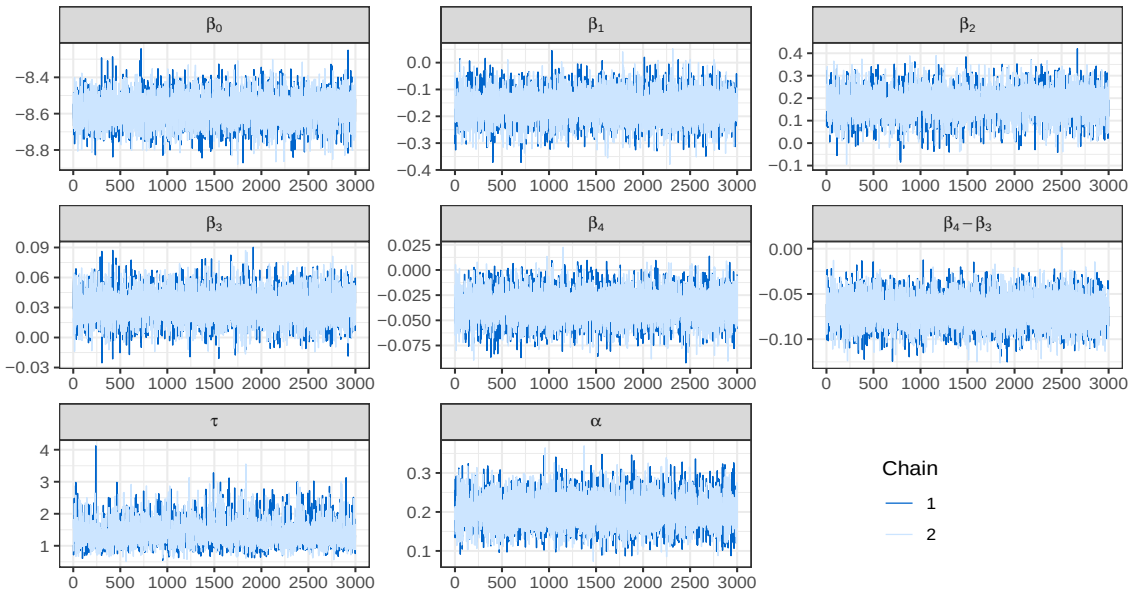

Figure D.4: Trace Plots for the posteriors of parameters without measurement error adjustment (under prior1).

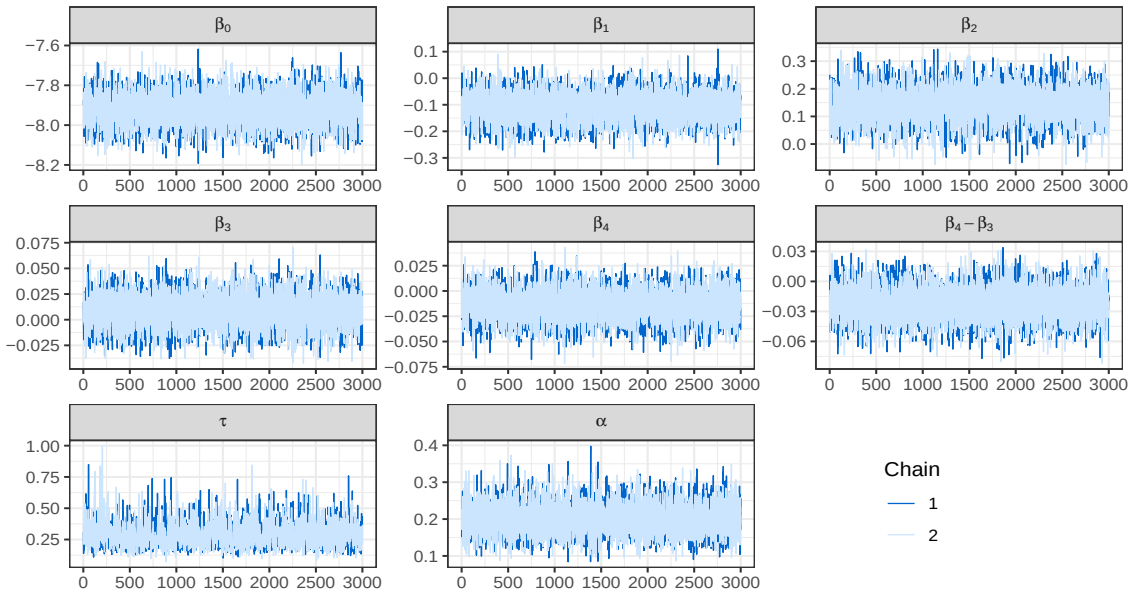

Figure D.5: Trace Plots for the posteriors of parameters with measurement error adjustment (under prior1).

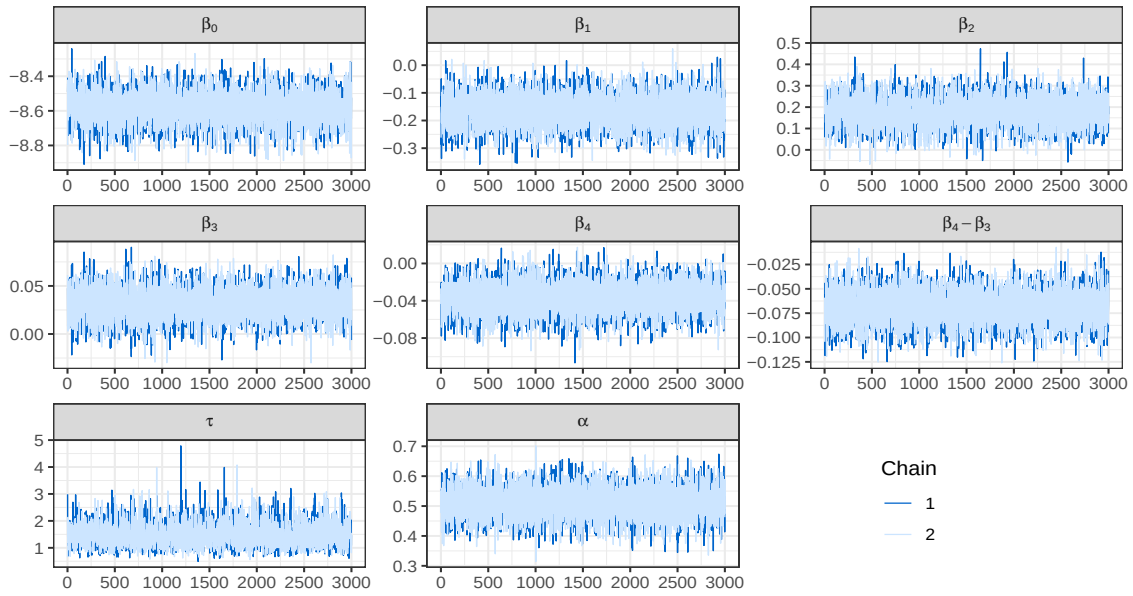

Figure D.6: Trace Plots for the posteriors of parameters without measurement error adjustment (under prior2).

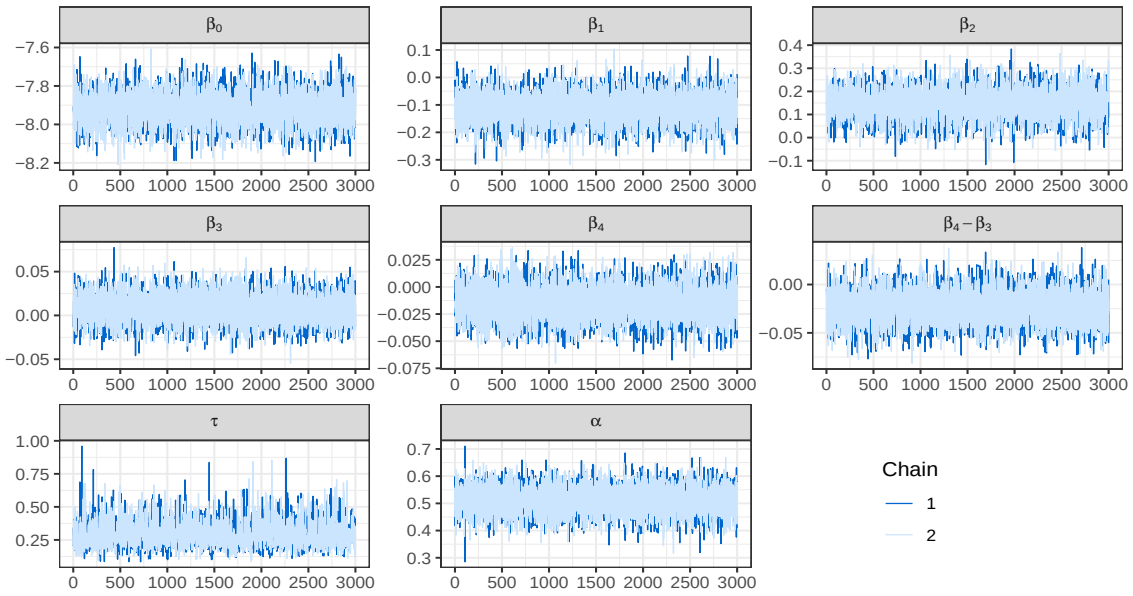

Figure D.7: Trace Plots for the posteriors of parameters with measurement error adjustment (under prior2).

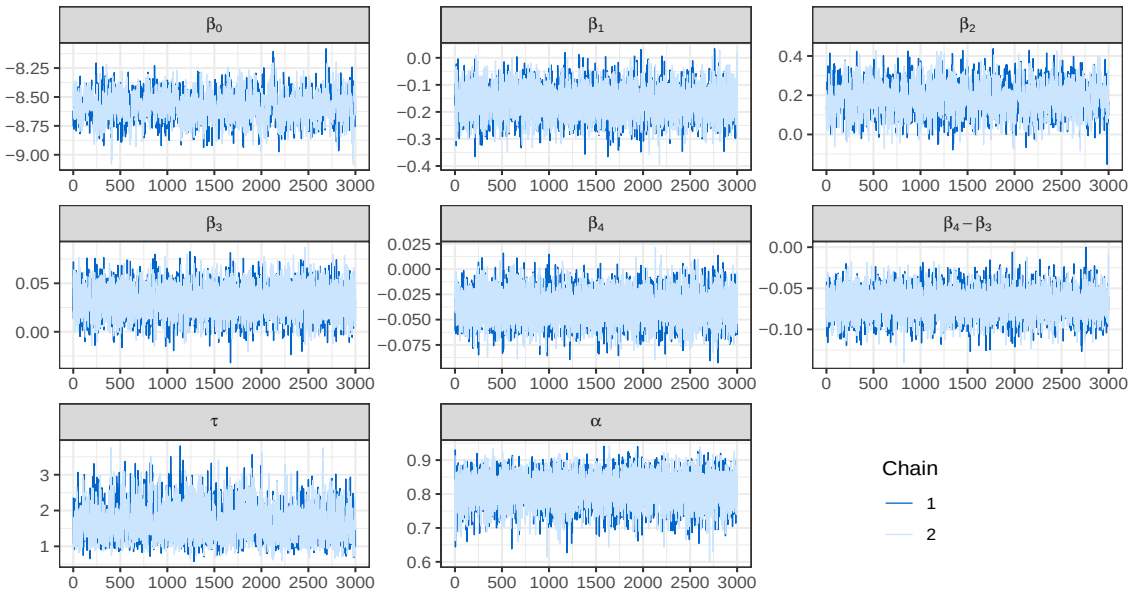

Figure D.8: Trace Plots for the posteriors of parameters without measurement error adjustment (under prior3).

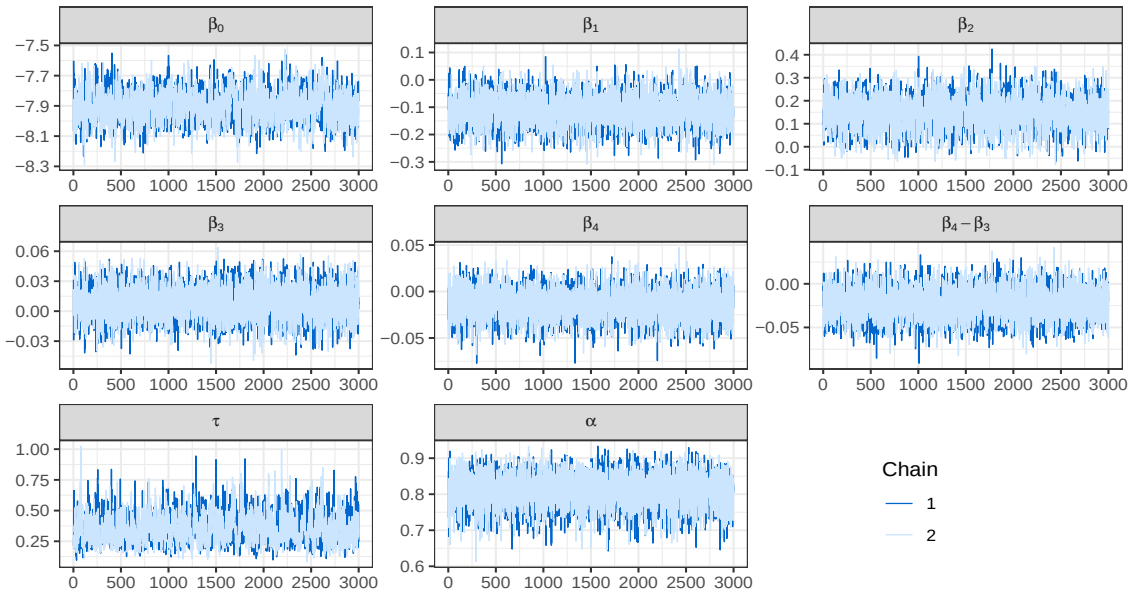

Figure D.9: Trace Plots for the posteriors of parameters with measurement error adjustment (under prior3).